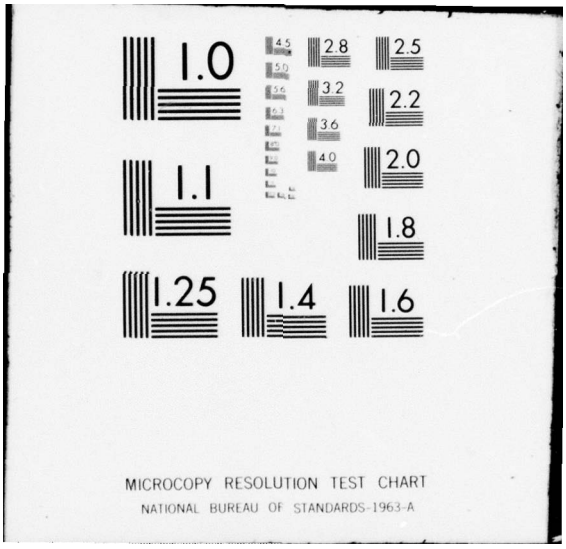
AD-A062 468 OHIO STATE UNIV COLUMBUS DEPT OF GEODETIC SCIENCE F/G 8/5
APPROXIMATION OF COVARIANCE FUNCTIONS BY NON-POSITIVE DEFINITE --ETC(U)
MAY 78 H SUENKEL F19628-76-C-0010
UNCLASSIFIED DGS-271 AFGL-TR-78-0177 NL

END
DATE FILMED
3-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFGL-TR-78-0177

(12) LEVEL II

APPROXIMATION OF COVARIANCE FUNCTIONS BY NON-POSITIVE DEFINITE FUNCTIONS

Hans Sünkel

The Ohio State University
Research Foundation
Columbus, Ohio 43212

May 1978

Scientific Report No. 15

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

(14) DGS-271, SCIENTIFIC-15

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (18) AFGL-TR-78-0177	2. GOVT ACCESSION NO. (9) Scientific rept.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) (6) APPROXIMATION OF COVARIANCE FUNCTIONS BY NON-POSITIVE DEFINITE FUNCTIONS .	5. TYPE OF REPORT & PERIOD COVERED Scientific. Interim. Scientific Report No. 15
	6. PERFORMING ORG. REPORT NUMBER Report No. 271
7. AUTHOR(s) (10) Hans Stinkel	8. CONTRACT OR GRANT NUMBER(s) (15) F19628-76-C-0010
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Geodetic Science The Ohio State University Columbus, Ohio 43210	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 62101F 760003AG (17) 03
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01731 Contract Monitor - Bela Szabo/LW	12. REPORT DATE (11) May 1978
	13. NUMBER OF PAGES 90
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 89p.	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DDC RECEIVED DEC 21 1978 B

18. SUPPLEMENTARY NOTES

TECH, OTHER

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Physical geodesy, Collocation, Covariance function, Approximation, Finite elements, Spline functions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In some applications of collocation we face two serious drawbacks, frequent calculations of linear functionals operating on the covariance function, and the inversion of a large matrix, both causing much computer time.

The frame of this work is an investigation how to avoid calculations of the exact covariance function and to replace it by some approximations. Three different kinds of approximating functions are studied, all of them being finite elements: the step function, the piecewise linear function and the cubic spline

next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

400 254

function.

After stating the essential properties of covariance functions, its approximations are discussed for the distance dependent covariance function and its extension into space.

FOREWORD

This report was prepared by Dr. Hans Sünkel, assistant to Dr. Helmut Moritz, Professor, Technical University at Graz and Adjunct Professor, Department of Geodetic Science of The Ohio State University, under Air Force Contract No. F19628-76-C-0010, The Ohio State University Research Foundation, Project No. 710334, Project Supervisor, Urho A. Uotila, Professor, Department of Geodetic Science. The contract covering this research is administered by the Air Force Geophysics Laboratory (AFGL), Hanscom Air Force Base, Massachusetts, with Mr. Bela Szabo, Project Scientist.

CONTENTS

1. Introduction

During the last decade, when least-squares collocation presented itself as the data processing model in physical geodesy, the most serious argument against was the inversion of a large matrix resulting in much computer time needed for this purpose. No attention was paid to the time used for calculating the linear functionals on the covariance function because for simple problems this time is definitely inferior to the inversion time. The situation, however, changed immediately when problems were attacked which involved many and/or difficult covariance calculations. Although R.H. Rapp and C.C. Tscherning succeeded in deriving closed expressions of covariance functions for different models of anomaly degree variances (Rapp and Tscherning, 1974), the closed expressions still consist of functions like logarithmic and trigonometric functions, which are expensive in terms of computer time. There is obviously no way out of this dilemma.

We mention only a few kinds of application: prediction of mean gravity anomalies over rectangular blocks from point values, prediction of mean gravity anomalies over larger areas from mean gravity anomalies over smaller areas, prediction of mean gravity anomalies from satellite altimetry data, all problems involving satellite dynamics. All these applications have one common feature: it is necessary to calculate covariances by numerical integration. In case of mean gravity anomaly prediction the integration is at most twofold, in case of satellite dynamics, however, it is multifold. In the former case an explicit integration procedure can be avoided, if one replaces the rectangular area of integration by a circular one. The so-called smoothing operation is caused by an isotropic smoothing operator acting on the covariance function which itself is also isotropic. Therefore, the convolution of the smoothing operator with the covariance function corresponds to a product of the corresponding eigenvalues, which is naturally very simple. In order to obtain

closed expressions for the mean gravity anomaly function, however, a further artificial assumption has to be made: the eigenvalues of the smoothing operator have to be replaced by some other values (K.P. Schwarz, 1976). In problems involving satellite dynamics as satellite-to-satellite ranging probably the only way to calculate covariances is by numerical integration over some time interval. Using exact covariances for the integration procedure is extremely time consuming (J. Kryński, 1978).

These were the reasons why the question arose whether it is possible to use some more or less accurate approximations of the exact covariance functions; the approximating functions should be simple, easy to handle, accurate and should consume as little mass storage as possible. A first attempt was reported to the author by G. Lachapelle (personal communication) who replaced the different covariance functions by step functions, the simplest approximating finite elements. However, one has to be very careful when using such approximations because of the possible lost of positive definiteness of the covariance matrix.

A better approximation can probably be achieved by piecewise linear elements and, even more, by cubic spline functions. From our point of view, the spectrum of the approximating function relative to the exact one gives a very good idea of its usefulness. Chapter 3-5, therefore, deals with the spectral properties of the functions considered here.

Due to the approximation of the basic covariance function the covariance matrix differs slightly from the exact one. This fact, again, will falsify to some extent the predicted signal as well as its error. It is not a simple task to estimate the consequences of inaccurate covariances. An attempt in this direction is performed in chapter 8.

In chapter 9 we shall present a spatial approximation model which is based on a bicubic spline function. Comparisons of computer time used for the exact and approximating covariance function finish our investigations.

2. Basic properties of covariance functions

Since collocation is already very well known to geodesists, we will, only for the sake of completeness, summarize the essential properties of covariance functions. When we speak about covariance functions here we always have the covariance function of the disturbing potential in mind.

The general form of an homogeneous isotropic spatial covariance function is given by

$$K(P,Q) = \sum_{n=0}^{\infty} k_n \left(\frac{R_b^2}{r_P r_Q}\right)^{n+1} P_n(\cos\psi) \quad , \qquad (2\text{-}1)$$

where
- P, Q ..... points in space,
- r_P, r_Q ... modulus of the radius vectors of P and Q ,
- R_b ... radius of the Bjerhammar sphere,
- ψ ... spherical distance between P and Q ,
- $P_n(\cos\psi)$.. Legendre polynomial of degree n ,
- k_n ... positive coefficients.

Because of the product $r_P r_Q$, $K(P,Q)$ is <u>symmetric</u> with respect to P and Q

$$K(P,Q) = K(Q,P) \quad ,$$

and moreover, the solid spherical harmonics guarantee <u>harmonicity</u> with respect to both points outside the sphere $r = R_b$

$$\Delta_P K(P,Q) = \Delta_Q K(P,Q) = 0 \ .$$

One important fact, to which no attention is paid sometimes, is the dependence on essentially two variables only: the spherical distance ψ and the product $r_P r_Q$. This property will be used extensively in chapter 9 when we approximate the spatial covariance function by a bicubic spline function dependent on these two variables.

According to the Funck-Hecke formula (Meissl, 1971, p.38) the spherical harmonics are the eigenfunctions of an integral transformation with a distance dependent kernel $J(P.Q)$ with eigenvalues λ_n :

$$\int_\sigma J(P.Q)\begin{Bmatrix} R_{nm}(Q) \\ S_{nm}(Q) \end{Bmatrix} d\sigma(Q) = \lambda_n \begin{Bmatrix} R_{nm}(P) \\ S_{nm}(P) \end{Bmatrix} \qquad (2\text{-}2a)$$

$$\lambda_n = 2\pi \int_{-1}^{1} J(t)P_n(t)dt \ , \quad t := \cos\psi \ . \qquad (2\text{-}2b)$$

Taking P and Q as points on the unit sphere, the eigenvalues of the covariance function (2-1) are given by

$$\lambda_n = 2\pi \sum_{j=0}^{\infty} k_j \int_{-1}^{1} P_j(t)P_n(t)dt \ , \qquad (2\text{-}3)$$

which, due to the mutual orthogonality of P_n , reduces to

$$\lambda_n = \frac{4\pi}{2n+1} k_n \ . \qquad (2\text{-}3)'$$

When all coefficients k_n , as assumed at the beginning, are positive, then also all eigenvalues λ_n will be positive. A positive spectrum (which is discrete in the spherical case and

consists of the eigenvalues λ_n), again, is a necessary and sufficient condition for the positive defineteness of the covariance function.

Similar relations as in the spherical case hold also for its planar analogue: A homogeneous, isotropic covariance function which is symmetric with respect to P and Q and harmonic in the upper half space $z \geq 0$ is given by

$$K(P,Q) = \int_0^\infty J_o(\eta s) G(\eta) e^{-\eta(z_P+z_Q)} \eta d\eta , \qquad (2\text{-}4)$$

with $J_o(x)$ ... Bessel function of first kind and zero order,
z_P, z_Q ... heights of P and Q above the plane $z = 0$,
s ... planar distance between P and Q ,
η ... frequency

and the continuous spectrum

$$G(\eta) = \int_0^\infty J_o(\eta s) K(s) s ds \qquad (2\text{-}5)$$

for $z_P = z_Q = 0$. $G(\eta)$ in equation (2-5) is the Hankel transform of $K(s)$, $K(s)$ in (2-4) is its inverse transform (apart from the spatial dependence expressed by z_P and z_Q). As in the spherical case, the positivity of the spectrum $G(\eta)$ is a necessary and sufficient condition for the positive definiteness of the corresponding covariance function.

It should be mentioned that all relations written down in this chapter are not new at all, but only an outline of (Moritz, 1976). However, for reasons of continuity, it seemed to be worth to write down them here again.

3. Approximation of the distance dependent covariance function by step functions

All finite elements share two essential properties: they have a compact support and are there quadratically integrable. The simplest finite element is the step function which is constant over some domain and vanishes outside. A function consisting of step functions is therefore, apart from the trivial case, discontinuous and as such an element out of $C_{-1}[D]$, the space of all discontinuous functions defined on the domain D . However, if the function is quadratically integrable, the space $C_{-1}[D]$ can be enlarged to $K_o[D]$, the space of all quadratically integrable functions which must not necessarily be continuous.

3.1 Step function approximation

Here we deal with an approximation of the covariance function by a step function which is naturally an element of $K_o[-1,1]$, the domain of definition being the closed interval $[-1,1]$ corresponding to the range of the cosine function. (We could also have chosen the domain $[0,\pi]$; however, for further applications the dependence on the variable $t = \cos\psi$ is more advantageous.

In order to make such an approximation we have first to arrange some gridding of the interval. It is evident that there exists an infinite number of such interval subdivisions. We select the most natural one with constant grid spacing equal to h . Let J be the number of subintervals within the total interval $[-1,1]$, then the grid points are given by

$$t_j = -1 + (j-1)h \ , \quad j = 1,\ldots, J+1 \qquad (3\text{-}1a)$$

with $h = 2/J$.

Now, the step function approximation is performed in such a way that the step height at the grid point t_j (which is the midpoint of the step) is equal to the value of the covariance functior at this point. Therefore, the set of function values, corresponding to the grid defined in (3-1), is given by

$$f_j = K(t_j) , \quad j = 1,\ldots, J + 1 . \tag{3-1b}$$

Figure 3.1 illustrates the approximation.

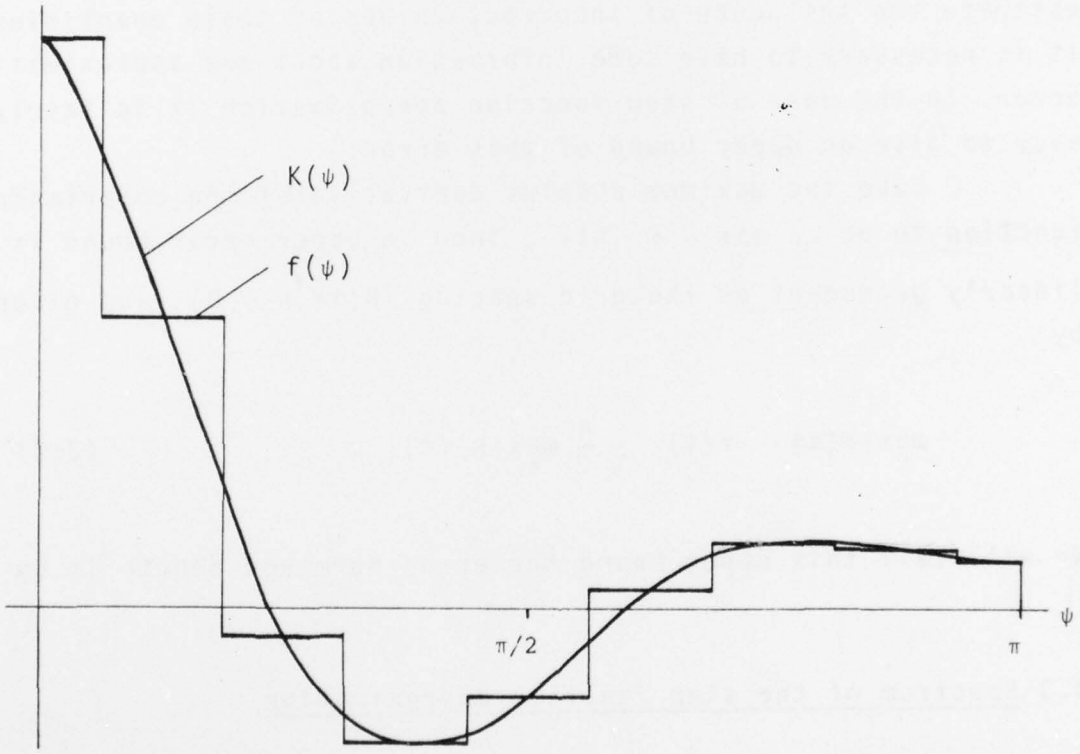


Fig. 3.1 Step function approximation

As stated above, the step function is discontinuous and therefore not differentiable. Consequently, such an approximation can only be used for one single kind of covariances. For example, separate approximating step functions have to be

used for the auto-covariance of the disturbing potential $\mathrm{cov}(T_P, T_Q)$ and for the cross-covariance between the disturbing potential and the deviation of the vertical $\mathrm{cov}(T_P, \xi_Q)$. This is certainly an essential disadvantage. An advantage, however, is the very simplicity of its expression.

3.2 Maximum approximation error

The estimate of the signal and its error depends on the different auto- and cross-covariances involved. In order to estimate the influence of incorrect values of these quantities, it is necessary to have some information about the approximation error. In the case of step function approximation it is fairly easy to give an upper bound of this error.

Take the maximum absolut derivative of the covariance function to be $\max_t |K'(t)|$, then an upper error bound is linearly dependent on the grid spacing h (if $h \to 0$) and given by

$$\max_t |K(t) - f(t)| \leq \frac{h}{2} \max_t |K'(t)| \quad . \tag{3-2}$$

We will call this upper bound the error norm and denote it by $\| \ \|$.

3.3 Spectrum of the step function approximation

As we have already stated in the introduction, the deviation of the approximation function from the exact one gives a good idea about the intrinsic properties of the approximation and therefore about its usefulness.

Taking equations (2-2b) and (3-1) into account, the eigenvalues of the step function approximation can be calculated from

$$\lambda_n = 2\pi \sum_{j=1}^{J+1} f_j \int_{t_j}^{t_{j+1}} P_n(t)dt \ . \qquad (3\text{-}3)$$

The integrals occurring here can be expressed in a closed form (Meissl, 1971, p.46):

$$\int_{x_j}^{x_{j+1}} P_n(t)dt = \frac{1}{2n+1}\left[P_{n+1}(x_{j+1}) - P_{n-1}(x_{j+1}) + P_{n-1}(x_j) - P_{n+1}(x_j)\right] ,$$

together with $\int_{x_j}^{x_{j+1}} P_o(t)dt = x_{j+1} - x_j$

and $\int_{x_j}^{x_{j+1}} P_1(t)dt = \frac{1}{2}(x_{j+1}^2 - x_j^2)$. $\qquad (3\text{-}4)$

A very simple example will demonstrate the principle: we choose $J = 1$, so that the approximating step function will have the form

$$f(t) = \begin{cases} f(1) & \text{for} \quad 0 \leq t \leq 1 \\ \\ f(-1) & \text{for} \ -1 \leq t < 0 \end{cases} .$$

According to (3-3) the eigenvalues of this simple model are given by

$$\lambda_n = 2\pi\left[f(-1)\int_{-1}^{0} P_n(t)dt + f(1)\int_{0}^{1} P_n(t)dt\right] .$$

With (3-4) we obtain

$$\lambda_n = 2\pi\{f(-1)\frac{1}{2n+1}\left[P_{n+1}(0) - P_{n-1}(0) + P_{n-1}(-1) - P_{n+1}(-1)\right]$$

$$+ f(1)\frac{1}{2n+1}\left[P_{n+1}(1) - P_{n-1}(1) + P_{n-1}(0) - P_{n+1}(0)\right]\} .$$

Let us now distinguish between even and odd values for n . For n even, the following relations hold:

n even: $P_{n+1}(0) = P_{n-1}(0) = 0$

$$P_{n+1}(-1) = P_{n-1}(-1) = -1$$

$$P_{n+1}(1) = P_{n-1}(1) = 1 .$$

Therefore, all eigenvalues λ_n with $n > 0$ and even vanish. This fact would, until now, at most allow positive semi-definiteness of the step function, presumed all odd eigenvalues are non-negative. Now we have still to investigate the eigenvalues of odd degree. For n odd we find:

n odd: $P_{n+1}(0) = -\frac{n}{n+1} P_{n-1}(0) \neq 0$

$$P_{n+1}(-1) = P_{n-1}(-1) = 1$$

$$P_{n+1}(1) = P_{n-1}(1) = 1 .$$

Consequently, we obtain for the eigenvalues of odd degree

$$\lambda_n = \frac{2\pi}{n+1}\left[f(1) - f(-1)\right]P_{n-1}(0) .$$

The expression in brackets is positive because of the fact, that the covariance is always smaller than the variance, $K(\psi) < K(0)$, $\psi \neq 0$. Therefore, the sign of λ_n depends on the sign of $P_{n-1}(0)$. It is not difficult to give a formula for $P_{n-1}(0)$:

$$P_{n-1}(0) = (-1)^{\frac{n-1}{2}} \frac{(n-1)!}{\left[\left(\frac{n-1}{2}\right)!\right]^2 2^{n-1}} ,$$

so that the λ_n finally are given by

$$\lambda_n = \begin{cases} 2\pi\left[K(0)-K(\pi)\right](-1)^{\frac{n-1}{2}} \dfrac{(n-1)!}{(n+1)\left[\left(\frac{n-1}{2}\right)!\right]^2 2^{n-1}} & \text{for } n \text{ odd} \\ 0 \quad \text{for } n \text{ even} . \end{cases}$$

It is obvious that the series of odd degree λ_n has alternating sign. Therefore, this simple approximation function is not even positive semi-definite. Figure 3.2 demonstrates the behaviour of the eigenvalues for $2\pi\left[K(0)-K(\pi)\right] = 1$. (The eigenvalue $\lambda_o = 2\pi\left[K(0)+K(\pi)\right]$.)

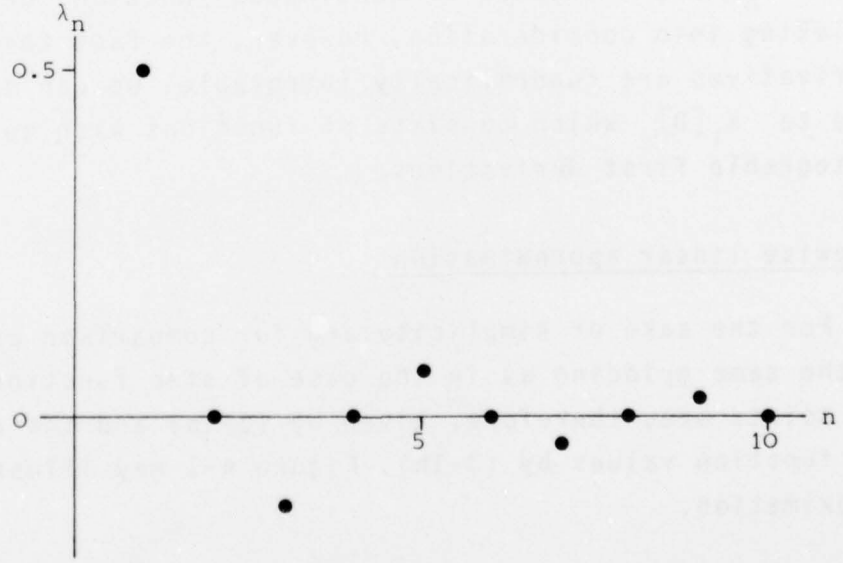


Fig. 3.2 Eigenvalues of a 2-step function

This example should only explain the principle; the bigger J (smaller h), the higher the n_o of the first non-positive eigenvalue will be and the less will the eigenvalues of the approximating function differ from the exact ones for $n < n_o$. Some realistic examples are given in chapter 6.

The step function approximation of the covariance function, however, will never be positive definite for $h > 0$. Therefore, one has to be extremely careful when using such approximations.

4. Approximation of the distance dependent covariance function by piecewise linear functions

A better approximation of the covariance function can be achieved by using a piecewise linear function. Such a function does not have discontinuities as the step function, but is continuous over the whole domain D of definition. It is, however, not continuously differentiable, and consequently an element of $C_o[D]$, the space of continuous functions defined on D. Taking into consideration, however, the fact that the first derivatives are quadratically integrable, we can enlarge the space to $K_1[D]$ which consists of functions with quadratically integrable first derivatives.

4.1 Piecewise linear approximation

For the sake of simplicity and for comparison purposes we take the same gridding as in the case of step functions. The grid points are, therefore, given by (3-1a) and the corresponding function values by (3-1b). Figure 4-1 may illustrate the approximation.

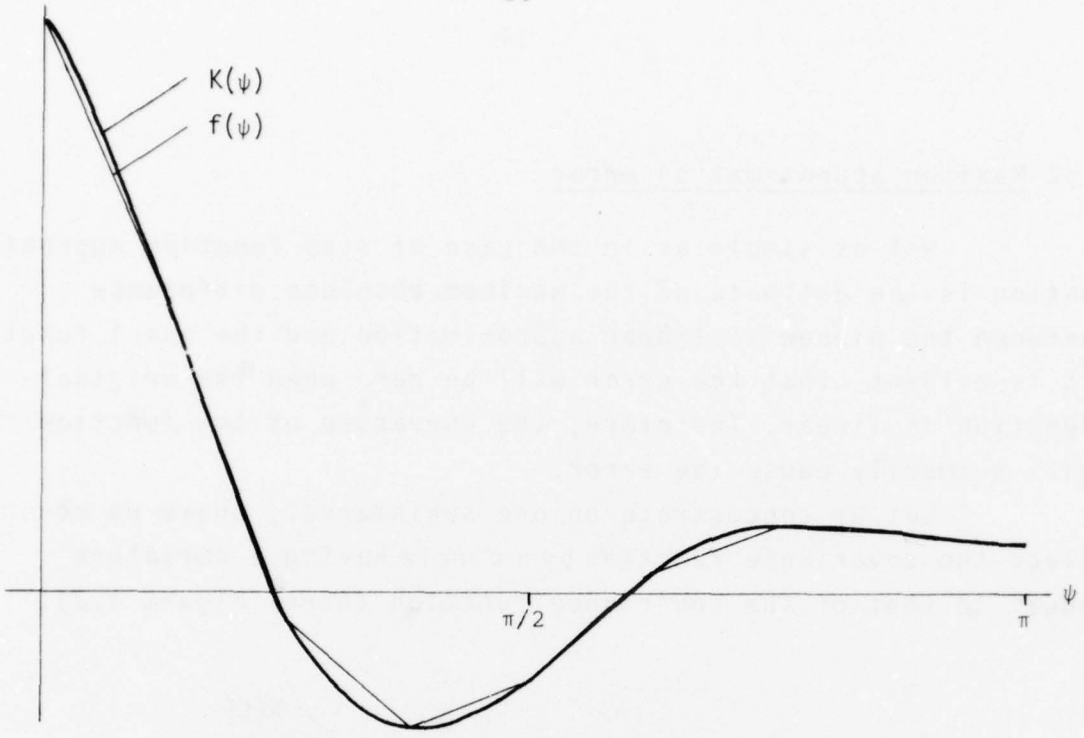


Fig. 4.1 Piecewise linear approximation

As stated above, the piecewiselinear function is continuous. Its derivative is a step function. Consequently, such an approximation can not only be used for one kind of covariances only, but also for such ones which involve at most one derivative with respect to t. For example, approximating the basic auto-covariance function of the disturbing potential $\mathrm{cov}(T_P,T_Q)$ by the piecewise linear function, the cross-covariance between the disturbing potential and the deviation of the vertical $\mathrm{cov}(T_P,\xi_Q)$ can also be derived by differentiation and is therefore a step function. Therefore, the area of application is already enlarged relative to the step function approach.

4.2 Maximum approximation error

Not as simple as in the case of step function approximation is the estimate of the maximum absolute difference between the piecewise linear approximation and the exact function. It is evident that the error will be zero when the original function is linear. Therefore, the curvature of the function will primarily cause the error.

Let us concentrate on one subinterval, where we replace the covariance function by a circle having a curvature equal to that of the covariance function there (Figure 4.2).

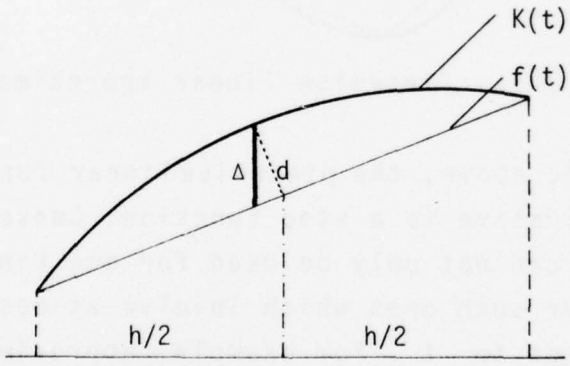


Fig. 4.2 Approximation element

The radius of curvature ρ of $K(t)$ is given by the well known formula

$$\rho = \frac{\{1 + [K'(t)]^2\}^{3/2}}{|K''(t)|} \quad . \tag{4-1}$$

With (4-1) it is easy to calculate d :

$$d = \rho - \sqrt{\rho^2 - \frac{h^2}{4}(1+K'^2)}$$

and with $\Delta = d\sqrt{1+K'^2}$ we obtain the error

$$\Delta = \frac{(1+K'^2)^2}{K''} \left(1 - \sqrt{1 - \left[\frac{hK''}{2(1+K'^2)}\right]^2}\right) . \qquad (4\text{-}2)$$

Consequently, an upper bound of the error can be estimated by

$$\max_t |K(t) - f(t)| \leq \max_t \left\{\frac{[1+K'(t)^2]^2}{K''(t)} \left(1 - \sqrt{1 - \left[\frac{hK''(t)}{2(1+K'(t)^2)}\right]^2}\right)\right\} .$$

If the product hK'' is small, we can evaluate the square root into a Taylor series and neglect higher terms obtaining

$$1 - \sqrt{1 - \left[\frac{hK''}{2(1+K'^2)}\right]^2} = \frac{h^2K''^2}{8(1+K'^2)^2} + O(h^4K''^4) .$$

Therefore, the error norm for sufficiently small $hK''(t)$ is given by the simple expression

$$\max_t |K(t) - f(t)| \leq \frac{h^2}{8} \max_t |K''(t)| . \qquad (4\text{-}3)$$

The same result, derived differently, can be found in (Strang & Fix, 1973. p.44).

Let us now estimate the error of the first derivative. Since $f(t)$ interpolates $K(t)$, the difference $K(t) - f(t)$ is zero at each grid point. Therefore, within each subinterval there must be at least one point t' where $K'(t') - f'(t') = 0$. Considering that $f(t)$ is linear, $K''(t) - f''(t) = K''(t)$ within each subinterval and we obtain the estimate

$$\max_t |K'(t) - f'(t)| \leq h \max_t |K''(t)| . \qquad (4\text{-}4)$$

A more careful proof of (4-4) would improve the dependence on h to h/2.

4.3 Spectrum of the piecewise linear approximation

The piecewise linear function on the subinterval $[t_j, t_{j+1}]$ is given by

$$f^{(j)}(t) = \sum_{\ell=0}^{1} a_\ell^{(j)} (t-t_j)^\ell = a_o^{(j)} + a_1^{(j)}(t-t_j) \tag{4-5}$$

with coefficients $a_o^{(j)} = K(t_j)$

and $a_1^{(j)} = \frac{1}{h}[K(t_{j+1}) - K(t_j)]$.

On introducing these relations into (2-2b) we obtain

$$\lambda_n = 2\pi \sum_{j=1}^{J} \int_{t_j}^{t_{j+1}} \Big[\sum_{\ell=0}^{1} a_\ell^{(j)} (t-t_j)^\ell \Big] P_n(t)dt$$

$$= 2\pi \sum_{j=1}^{J} \Big[(a_o^{(j)} - a_1^{(j)} t_j) \int_{t_j}^{t_{j+1}} P_n(t)dt + a_1^{(j)} \int_{t_j}^{t_{j+1}} P_n(t)tdt \Big]. \tag{4-6}$$

The expression for the first integral has already been derived and is given by equation (3-4). In order to evaluate the second integral we obey from (Hobson, 1965,p.38) the general relation

$$\int_{t_o}^{1} P_n(t)P_m(t)dt = \frac{1}{(n-m)(n+m+1)} \big[nP_m(t_o)P_{n-1}(t_o) - mP_n(t_o)P_{m-1}(t_o) - t_o(n-m)P_n(t_o)P_m(t_o) \big] \tag{4-7}$$

which is valid for $n \neq m$. Taking into account the fact that $t = P_1(t)$, the second integral can be expressed by

$$\int_{t_j}^{t_{j+1}} P_n(t)tdt = \int_{t_j}^{1} P_n(t)tdt - \int_{t_{j+1}}^{1} P_n(t)tdt$$

$$= \frac{1}{n^2+n-2}\Big\{\big[\, nt_j P_{n-1}(t_j) - P_n(t_j) - t_j^2(n-1)P_n(t_j)\big]$$

$$- \big[\, nt_{j+1} P_{n-1}(t_{j+1}) - P_n(t_{j+1}) \qquad (4\text{-}8a)$$

$$- t_{j+1}^2(n-1)P_n(t_{j+1})\big]\Big\},$$

which holds for $n \neq 1$; for $n = 0$ we have

$$\int_{t_j}^{t_{j+1}} P_n(t)tdt = \frac{1}{2}(t_{j+1}^2 - t_j^2)$$

and for $n = 1$

$$\int_{t_j}^{t_{j+1}} P_n(t)tdt = \frac{1}{3}(t_{j+1}^3 - t_j^3) \; . \qquad (4\text{-}8b)$$

The following example may demonstrate the principle. For comparison purposes we choose the same subdivision as for the step function case, $J = 1$, so that the approximating linear function has the following form:

$$f(t) = a_o + a_1(t+1)$$

with $a_o = K(\pi)$ and $a_1 = \frac{1}{2}[K(0) - K(\pi)]$.

According to (4-6) and (4-8) the eigenvalues of this simple model are given by

$$\lambda_n = 2\pi\left[(a_o + a_1)\int_{-1}^{1} P_n(t)dt + a_1\int_{-1}^{1} P_n(t)tdt\right],$$

which, because of the mutual orthogonality of Legendre polynomials, gives the simple solution

$$\lambda_o = \pi\left[K(0) + K(\pi)\right]$$

$$\lambda_1 = \frac{2\pi}{3}\left[K(0) - K(\pi)\right]$$

$$\lambda_n = 0 \quad \text{for} \quad n \geq 2 \ .$$

This result seems to be astonishing at first sight. However, when we concentrate a little bit more on it we find it obvious: The linear function $f(t)$ defined on the closed interval is linear in t which means that in reality $f(t)$ is a linear combination of $P_o(\cos\psi)$ and $P_1(\cos\psi)$, the Legendre polynomials of zero and first degree. Therefore, there can only exist non-vanishing eigenvalues of degree zero and one in agreement with our result. The approximated covariance function, however, is only positive semi-definite. But this is an exceptional case; for $J > 1$ the function is not even positive semi-definite as will be shown numerically in chapter 6.

5. Approximation of the distance dependent covariance function by cubic spline functions

Spline functions are very useful tools for purposes of interpolation and approximation. They are widely applied in all areas where numerical mathematics enters somehow. The theo-

retical background of spline functions is remarkably deep and extremely interesting especially because of its minimum norm and best approximation properties making interpolating spline functions unique among all other interpolating functions. We do not present details here; the interested reader may consult (Ahlberg et.al., 1967), the standard literature about spline functions. Some applications can be found in (Sünkel, 1977).

For the sake of completeness we give a short definition of a cubic spline: A function, twice continuously differentiable on the whole domain, connecting subsequent data points by cubic polynomials and fulfilling two special boundary conditions, is called an interpolating cubic spline. The coefficients of the cubic polynomials are uniquely determined by continuity conditions up to and including the second order derivative. Consequently, the cubic spline is an element of $C_2[D]$. For analogous reasons as in chapters 3 and 4 it can even be considered as an element of $K_3[D]$, the space of all continuous and twice continuously differentiable functions defined on D with quadratically integrable second order derivatives. Within each subinterval its equation is given by

$$f^{(j)}(t) = \sum_{\ell=0}^{3} a_\ell^{(j)} (t-t_j)^\ell$$

with

$$\begin{aligned} a_1^{(j)} &= f_j \\ a_2^{(j)} &= (f_{j+1}-f_j)/(x_{j+1}-x_j) - \frac{1}{6}(x_{j+1}-x_j)(f''_{j+1}+2f''_j) \\ a_3^{(j)} &= \frac{1}{2} f''_j \\ a_4^{(j)} &= \frac{1}{6}(f''_{j+1}-f''_j)/(x_{j+1}-x_j) \ . \end{aligned} \qquad (5\text{-}1)$$

The second derivatives $\{f''_j\}$ at the grid points $j=1,\ldots,J+1$ are derived from continuity conditions of first order derivatives.

5.1 Cubic spline approximation

Again we adopt the same gridding as in the step and linear case. The grid points are, therefore, given by (3-1a) and the corresponding function values by (3-1b). In order to give an idea about the performance of spline function approximation we have divided the whole interval [-1,1] into 12 subintervals only and have fitted the cubic spline to the covariances at these grid points. Figure 5.1 shows both curves, the exact covariance function and its spline approximation. The naked eye is not able to find a difference between both curves.

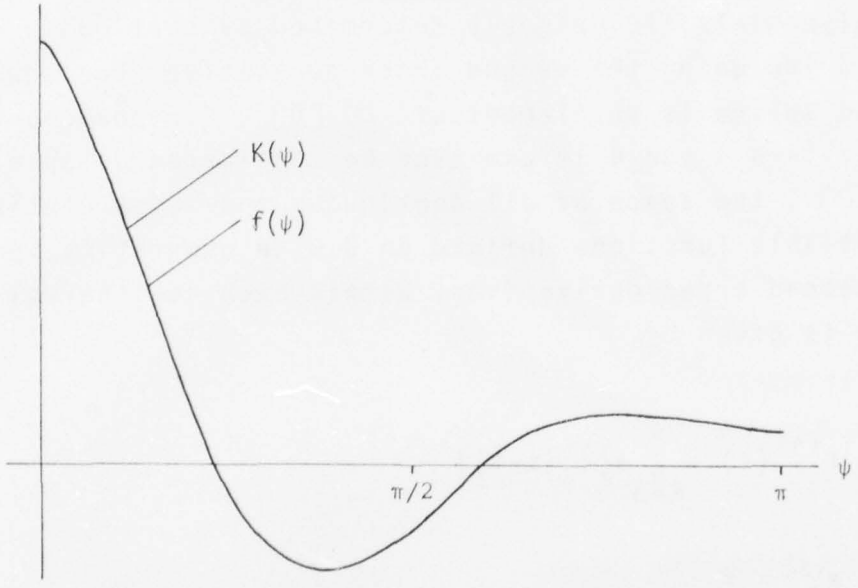


Fig. 5.1 Cubic spline approximation

As stated above, the cubic spline is twice continuously differentiable. Its first derivative is, therefore, a polynomial of second degree which is once continuously differentiable; its second derivative is a piecewise linear function being not continuously differentiable; finally, its third order derivative is a step function being discontinuous. Consequently, such a spline approximation can not only be used for one kind of covariances as this is the case for the step function approximation

or for covariances involving one derivative with respect to t, but for all covariances resulting from at most three differentiations with respect to the argument t. For example, approximating the basic auto-covariance function of the disturbing potential $\mathrm{cov}(T_P, T_Q)$ by a cubic spline, the cross-covariances between

$$(T_P, \ell_Q)\ ,\ (T_P, \frac{\partial^2 T_Q}{\partial \Theta^2})\ ,\ (T_P, \frac{\partial^2 T_Q}{\partial \Theta \partial \lambda})\ ,\ (T_P, \frac{\partial^2 T_Q}{\partial \lambda^2})\ ,\ (\ell_P, \frac{\partial^2 T_Q}{\partial \Theta^2})\ ,$$

$$(\ell_P, \frac{\partial^2 T_Q}{\partial \Theta \partial \lambda})\ ,\ (\ell_P, \frac{\partial^2 T_Q}{\partial \lambda^2}) \quad \text{and also the auto-covariances} \quad (\ell_P, \ell_Q)\ ,$$

with ℓ denoting the deviation of the vertical ξ and η, respectively, can be derived. The area of application is, therefore, much larger than in the case of piecewise linear approximation.

5.2 Maximum approximation error

It is not a simple task to estimate the maximum approximation error for cubic spline functions. The estimation is strongly dependent on the continuity class of the function space of which the original function is an element. Here we have to approximate the covariance function which is infinitely often continuously differentiable. However, we shall relax this property and content ourselves with a lower function space.

In (Ahlberg et.al., 1967, p.19ff.) we find lengthy derivations of such maximum approximation errors for different classes of functions. We are not going to rederive the formulas given there and present only the result specified to our case:

When $K'''(t)$ satisfies a Hölder condition of order α $(0 < \alpha \leq 1)$ on $[-1,1]$

$$|K'''(t) - K'''(t')| \leq A|t-t'|^{\alpha}\ ,$$

which is for the covariance function trivially true for $\alpha = 1$ and $A = \max_t |K^{IV}(t)|$, then it can be shown that the approximation error of the third derivative can be estimated by

$$\max_t |K'''(t) - f'''(t)| \leq 51h \max_t |K^{IV}(t)| . \qquad (5\text{-}2a)$$

Applying the theorem of Rolle we can furthermore estimate the maximum approximation errors for lower order derivatives and the function itself:

$$\max_t |K''(t) - f''(t)| \leq \frac{5}{6} \cdot 51h^2 \max_t |K^{IV}(t)| , \qquad (5\text{-}2b)$$

$$\max_t |K'(t) - f'(t)| \leq \frac{5}{6} \cdot 51h^3 \max_t |K^{IV}(t)| , \qquad (5\text{-}2c)$$

$$\max_t |K(t) - f(t)| \leq \frac{5}{6} \cdot 51h^4 \max_t |K^{IV}(t)| . \qquad (5\text{-}2d)$$

It is quite interesting to compare this result with the step function and piecewise linear function estimates. The error of covariance function approximation depends on the fourth power of the grid spacing and linearly on the fourth derivative of the covariance function. Consequently, comparing (3-2), (4-3) and (5-2d) we can give the following rule of thumb: If the approximation function is an element of $K_\mu[-1,1]$, the maximum approximation error depends on the $(\mu+1)$th power of the grid spacing multiplied by the maximum absolute value of the $(\mu+1)$th order derivative of the covariance function. Each differentiation, if it exists, decreases the dependence to one power of the grid spacing.

5.3 The spectrum of the cubic spline approximation

The cubic spline function within each subinterval is given by equation (5-1). The Fourier coefficients are the result of the projection of the spline onto the set of Legendre polynomials. We introduce (5-1) into (2-2b) and obtain

$$\lambda_n = 2\pi \sum_{j=1}^{J} \int_{t_j}^{t_{j+1}} \left[\sum_{\ell=0}^{3} a_\ell^{(j)} (t-t_j)^\ell \right] P_n(t)dt \ . \qquad (5\text{-}3)$$

Expressing powers of t in terms of Legendre polynomials we get

$$\begin{aligned}\lambda_n = 2\pi \sum_{j=1}^{J} \{ & \int_{t_j}^{t_{j+1}} P_n(t)P_o(t)dt \left[a_o^{(j)} - a_1^{(j)} t_j + a_2^{(j)} t_j^2 - a_3^{(j)} t_j^3 + \frac{a_2^{(j)}}{3} - a_3^{(j)} t_j\right] \\ & + \int_{t_j}^{t_{j+1}} P_n(t)P_1(t)dt \left[a_1^{(j)} - 2a_2^{(j)} t_j + 3a_3^{(j)} t_j^2 + \frac{3}{5} a_3^{(j)}\right] \\ & + \int_{t_j}^{t_{j+1}} P_n(t)P_2(t)dt \left[\frac{2}{3} a_2^{(j)} - 2a_3^{(j)} t_j\right] \\ & + \int_{t_j}^{t_{j+1}} P_n(t)P_3(t)dt \cdot \frac{2}{5} a_3^{(j)} \ . \end{aligned} \qquad (5\text{-}3)'$$

In this formula all integrals for $n > 3$ can be calculated according to (4-7). The integrals with $n \leq 3$ have partly already been calculated. Summarizing we obtain:

$$\int_{t_j}^{t_{j+1}} P_o(t)P_o(t)dt = t_{j+1} - t_j$$

$$\int_{t_j}^{t_{j+1}} P_o(t)P_1(t)dt = \frac{1}{2}(t_{j+1}^2 - t_j^2)$$

$$\int_{t_j}^{t_{j+1}} P_o(t)P_2(t)dt = \frac{1}{2}(t_{j+1}^3 - t_j^3 - t_{j+1} + t_j)$$

$$\int_{t_j}^{t_{j+1}} P_o(t)P_3(t)dt = \frac{1}{2}\left[\frac{5}{4}(t_{j+1}^4 - t_j^4) - \frac{3}{2}(t_{j+1}^2 - t_j^2)\right]$$

$$\int_{t_j}^{t_{j+1}} P_1(t)P_1(t)dt = \frac{1}{3}(t_{j+1}^3 - t_j^3)$$

$$\int_{t_j}^{t_{j+1}} P_1(t)P_2(t)dt = \frac{1}{4}\left[\frac{3}{2}(t_{j+1}^4 - t_j^4) - (t_{j+1}^2 - t_j^2)\right]$$

$$\int_{t_j}^{t_{j+1}} P_1(t)P_3(t)dt = \frac{1}{2}\left[(t_{j+1}^5 - t_j^5) - (t_{j+1}^3 - t_j^3)\right] \qquad (5\text{-}4)$$

$$\int_{t_j}^{t_{j+1}} P_2(t)P_2(t)dt = \frac{1}{4}\left[\frac{9}{5}(t_{j+1}^5 - t_j^5) - 2(t_{j+1}^3 - t_j^3) + (t_{j+1} - t_j)\right]$$

$$\int_{t_j}^{t_{j+1}} P_2(t)P_3(t)dt = \frac{1}{8}\left[5(t_{j+1}^6 - t_j^6) - 7(t_{j+1}^4 - t_j^4) + 3(t_{j+1}^2 - t_j^2)\right]$$

$$\int_{t_j}^{t_{j+1}} P_3(t)P_3(t)dt = \frac{1}{4}\left[\frac{25}{7}(t_{j+1}^7 - t_j^7) - 6(t_{j+1}^5 - t_j^5) + 3(t_{j+1}^3 - t_j^3)\right].$$

As in chapters 3 and 4 we also give here a simple example. We choose $J = 1$ and approximate the covariance function on the whole interval $[-1,1]$ by a single polynomial of third degree. In order to determine its four coefficients uniquely we need, apart from the function values at the end-points of the interval, two additional informations about the polynomial.

Since the polynomial is of third degree in t, it is a linear combination of Legendre polynomials $P_o(t)$, $P_1(t)$, $P_2(t)$ and $P_3(t)$. The Fourier coefficients of this approximation result from the polynomial's projection onto the set

$\{P_n(t)\}$, $n = 0,1,\ldots$. Therefore, we can already anticipate the behaviour of the eigenvalues: the eigenvalues up to and including degree 3 will in general be different from zero, all others will vanish.

We approximate the covariance function by the cubic polynomial for two different boundary conditions; a) for first order derivatives prescribed and b) for second order derivatives given.

Let us now discuss polynomial (a):
The coefficients of the polynomial are uniquely determined by $K(1)$, $K(-1)$, $K'(1)$, $K'(-1)$. At this point we must be careful not to put $K'(1)$ and $K'(-1)$ equal to zero: $dK(\psi)/d\psi$ is zero at $\psi = 0$ and $\psi = \pi$, but $dK(t)/dt$ does not at all vanish at the corresponding points $t = -1$ and $t = 1$. Taking this fact into account we obtain the following coefficients:

$$a_o = \frac{1}{2}\left[K(1)+K(-1)\right] - \frac{1}{4}\left[K'(1)-K'(-1)\right]$$

$$a_1 = \frac{3}{4}\left[K(1)-K(-1)\right] - \frac{1}{4}\left[K'(1)+K'(-1)\right]$$

$$a_2 = \frac{1}{4}\left[K'(1)-K'(-1)\right]$$

$$a_3 = \frac{1}{4}\left[K'(1)+K'(-1)\right] - \frac{1}{4}\left[K(1)-K(-1)\right] .$$

The projection of the polynomial onto the Legendre polynomials gives the eigenvalues

$$\lambda_o = 2\pi\{\left[K(1)+K(-1)\right] - \frac{1}{3}\left[K'(1)-K'(-1)\right]\}$$

$$\lambda_1 = \frac{2\pi}{15}\{6\left[K(1)-K(-1)\right] - \left[K'(1)+K'(-1)\right]\}$$

$$\lambda_2 = \frac{2\pi}{15}\left[K'(1)-K'(-1)\right]$$

$$\lambda_3 = \frac{2\pi}{35}\{-\left[K(1)-K(-1)\right] + \left[K'(1) + K'(-1)\right]\} \ .$$

For the case of model 2 covariance function (Tscherning, 1976) we have calculated the approximating polynomial which is presented in Figure 5.2. By forcing the polynomial to reproduce the first derivative at $t = 1$, we notice right away the polynomials revenge: it produces much too big curvatures at $t = 1$ and, moreover, it dips into the negative so much that one of the most essential conditions of a covariance function $K(t) \leq K(1)$ is not satisfied anymore.

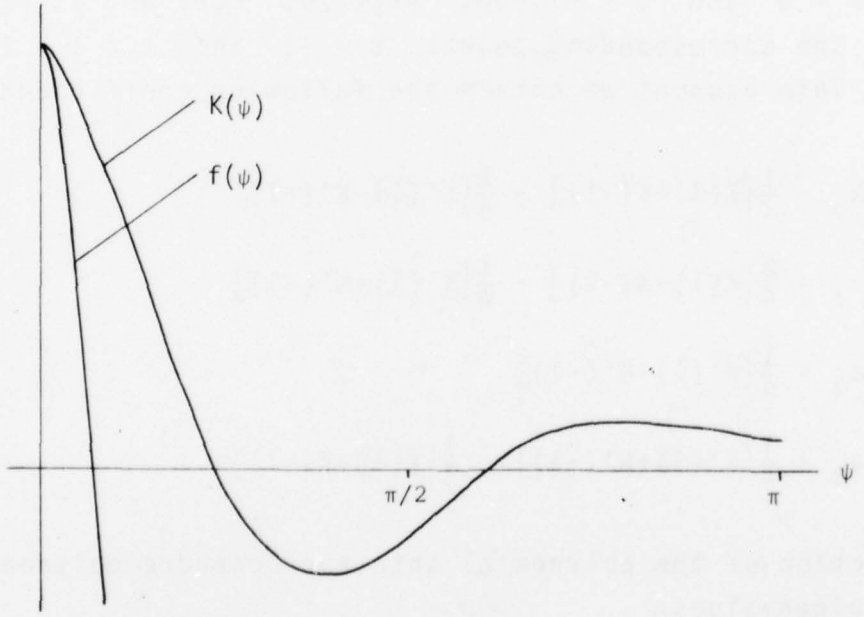


Fig. 5.2 Cubic polynomial approximation; boundary conditions: first derivatives at the endpoints

Polynomial (b) based on function values and second derivatives at the endpoints has even wilder features. The reason is that $D_t^2 K\big|_{t=1}$ is very big; therefore, it has a strong influence and controls more or less the behaviour of the polynomial over the whole domain. The coefficients are given by

$$a_o = \frac{1}{2}[K(1)+K(-1)] - \frac{1}{4}[K''(1)+K''(-1)]$$

$$a_1 = \frac{1}{2}[K(1)-K(-1)] - \frac{1}{12}[K''(1)-K''(-1)]$$

$$a_2 = \frac{1}{4}[K''(1)+K''(-1)]$$

$$a_3 = \frac{1}{12}[K''(1)-K''(-1)] \ ,$$

for the eigenvalues we obtain

$$\lambda_o = 2\pi\{[K(1)+K(-1)] - \frac{1}{3}[K''(1)+K''(-1)]\}$$

$$\lambda_1 = \frac{2\pi}{3}\{[K(1)-K(-1)] - \frac{1}{15}[K''(1)-K''(-1)]\}$$

$$\lambda_2 = \frac{2\pi}{15}[K''(1)+K''(-1)]$$

$$\lambda_3 = \frac{2\pi}{105}[K''(1)-K''(-1)] \ .$$

Its behaviour is demonstrated in Figure 5.3 (note the scale).

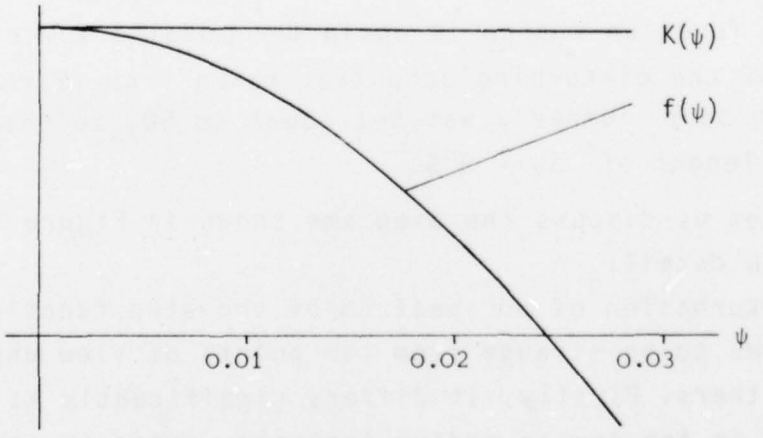


Fig. 5.3 Cubic polynomial approximation; boundary conditions: second derivatives at the endpoints

It should be pointed out once more that all these simple approximations to the true covariance function considered were only presented in order to give an idea about its features; nobody, however, is invited to use them. In any case: choose J big!

6. Perturbation of the spectrum of the covariance function

Very important for all approximations is the question, how much the spectrum is disturbed relative to the exact one and what is the degree of the first vanishing or negative eigenvalue. Evidently, the eigenvalue behaviour depends somehow on the degree of approximation. For comparison purposes we give the relative eigenvalue perturbation q_n defined by

$$q_n := \left| 1 - \frac{\tilde{\lambda}_n}{\lambda_n} \right|$$

with λ_n ... exact eigenvalue,
$\tilde{\lambda}_n$ ... eigenvalue of the approximating function

for the approximation models considered here (Figure 5.4). The covariance function chosen is again the model 2 covariance function of the disturbing potential taken from (Tscherning, 1976). The subdivision number J was set equal to 50, so that each interval has a lenght of $\Delta\psi \doteq 3^{\circ}6$.

Let us discuss the diagrams shown in Figure 5.4 a little bit more in detail:

a) The perturbation of the spectrum of the step function approximation seems to be strange from two points of view when compared with the others. Firstly, it differs significantly as far as the high bulge in the low to medium frequency range is concerned and, secondly, it shows oscillations of high frequency over the whole range. Both phenomena are due to the simplicity of the

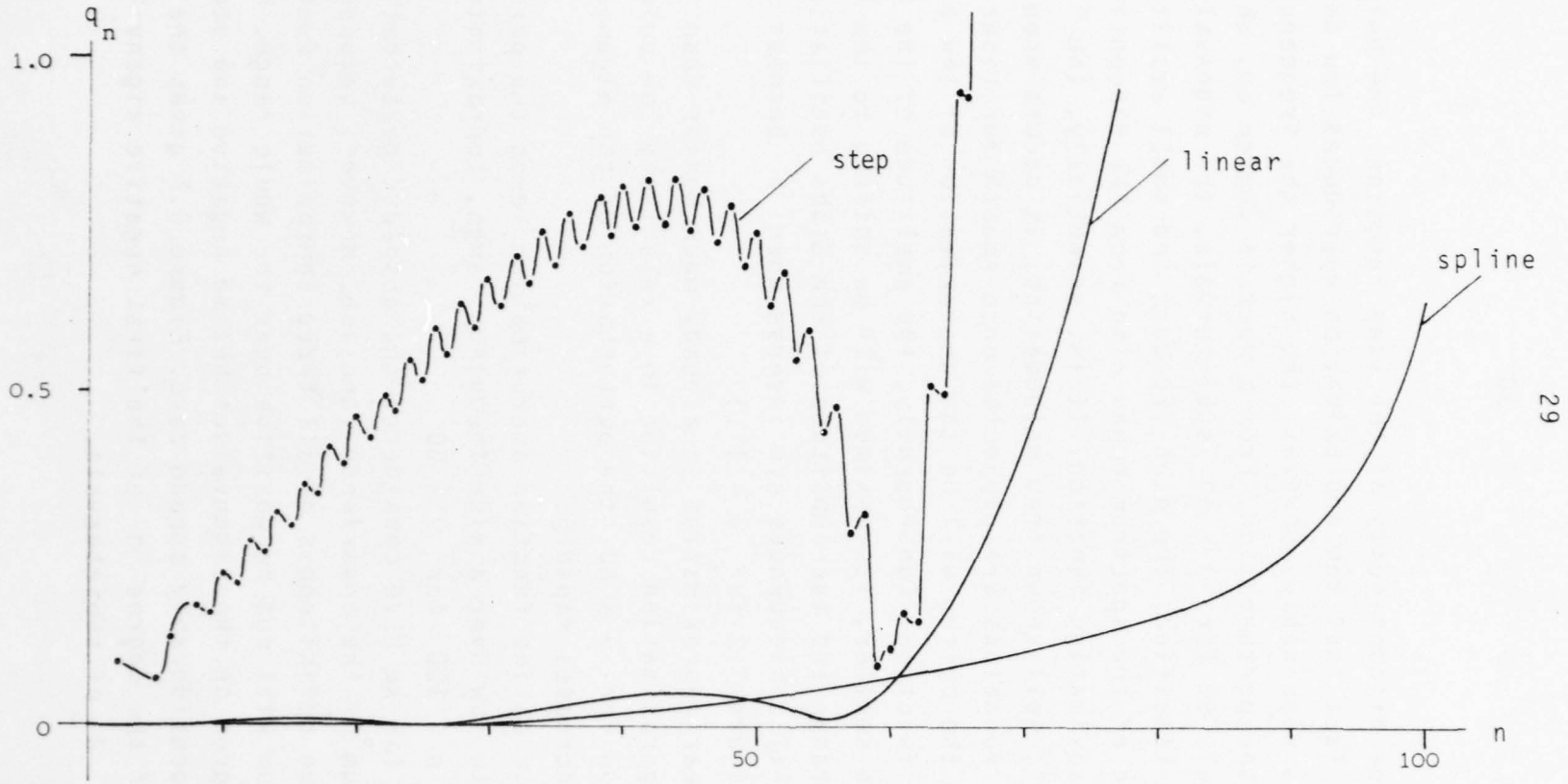


Fig. 5.4 Relative perturbation of the spectrum (spherical case, J = 50)

element, the discontinuity of the step function. The bulge reflects the fact that the approximation reproduces low degree eigenvalues reasonably, however, the higher the frequency the bigger is the perturbation. From a certain degree on, which is approximately 65 for $J = 50$ subintervals, the eigenvalues are completely falsified. The high-frequent and small amplitude oscillation of the spectrum comes also from the discontinuity of the approximating function. It is, essentially, the "Gibbs-phenomenon", well known from mathematics. It occurs when discontinuous functions are projected onto smooth functions. The higher J the better will be the approximation of the true covariance function. Consequently, the amplitude of the bulge will become smaller, its maximum will be shifted to the higher frequency range and the amplitude of the Gibbs oscillation will decrease. All consequences are inverse when J becomes smaller. (Compare Figure 3.2 for $J = 1$.)

b) The linear approximation is already much better than the step function approximation resulting in a relatively low bulge. From degree 75 on for $J = 50$ the perturbation of the eigenvalues, however, increases rapidly.

c) The cubic spline function approximation keeps the perturbation quite low over a wide frequency range, approximately up to degree $n = 100$ for $J = 50$.

So far we have considered the absolute perturbation of the spectrum of the covariance function. However, because of the non-positive definiteness of all three approximation functions the spectrum will not be positive over the whole range. From a certain degree on the eigenvalues become negative and start to oscillate considerably around zero. Figure 5.5 gives the dependence of the degree n of the first negative eigenvalue on the number J of subintervals.

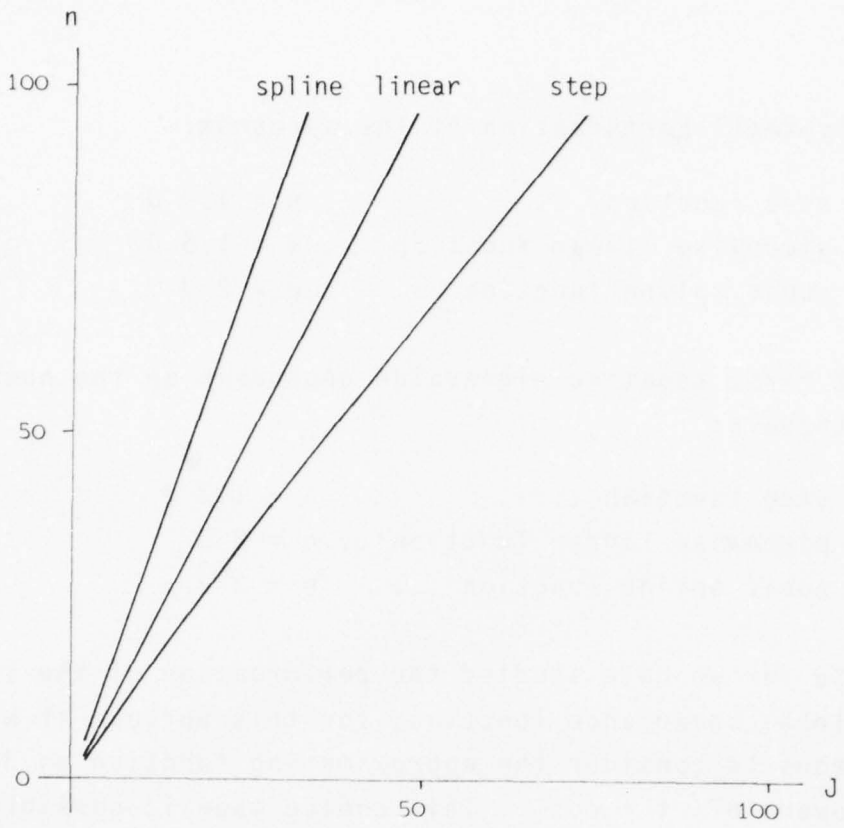


Fig. 5.5 Degree of the first negative eigenvalue dependent on the number J of subintervals

Again, we obey the quality of the spline approximation compared with the two others. It is quite interesting that the dependence is practically linear in all three cases.

Summarizing, the following conclusions can be drawn: The perturbation of the spectrum (of the eigenvalues, of the degree variance model) depends strongly on the smoothness of the approximating function and on the number J of subintervals chosen. Moreover, this dependence is almost linear (at least in the frequency range considered here). The following relations are rules of thumb and give an idea about the quality of approximations in the spectral domain.

Reasonably small perturbation of the spectrum:

step function ... $n = 1.3\ J$
piecewise linear function ... $n = 1.5\ J$
cubic spline function ... $n = 2\ J$.

Degree of first negative eigenvalue dependent on the number J of subintervals:

step function ... $n = 1.3\ J$
piecewise linear function ... $n = 2\ J$
cubic spline function ... $n = 3\ J$.

So far we have studied the perturbation of the spectrum of the global covariance function; for this purpose it was advantageous to consider the approximating function as dependent on the powers of $t = \cos\psi$. This choice made it possible to calculate the spectrum without numerical integration. However, for applications (especially for local ones) it is more appropriate to change the independent variable from t to ψ itself. The main advantages are:

1.) The function is now symmetric with respect to $\psi = 0$.
2.) A constant grid spacing corresponds to constant physical distances between gridpoints.(The mapping $t = \cos\psi$ maps an equispaced ψ-mesh onto a non-equispaced t-mesh, and inversely, the inverse mapping $\psi = \arccos t$ maps an equispaced t-mesh onto a non-eqispaced ψ-mesh which has disadvantages for practical applications.)
3.) Using a cubic spline repesentation of the covariance function, the spline subjected to the boundary conditions $D^2K(\psi)$ at the endpoints, because of its symmetry with respect to $\psi = 0$, can easily be made to have a vanishing first derivative $D\,K(\psi)$

at $\psi = 0$. This fact has very important consequences: The ψ-dependent spline can be defined in such a way that it exactly reproduces the variance, the vanishing first derivative at $\psi = 0$, the curvature at $\psi = 0$ and practically also the correlation length of the covariance function considered. These are precisely the essential parameters of a covariance function (Moritz, 1976) which control the prediction and, therefore, should not be changed by an approximation procedure.

Not all of these requirements are met by the t-spline which is, consequently, inferior to the ψ-spline. For these reasons we shall work in the sequel with the spline, dependent on the independent variable ψ . Nevertheless, it should be mentioned that the t-dependent spline as well as the ψ-dependent spline approximation can be made arbitrarily close to the exact covariance function when the grid spacing tends to zero.

7. Spectra of covariance function approximations in the plane

Many applications of collocation are local and not global. In such a case it is convenient to approximate the terrestrial sphere locally by a plane and to use planar equivalents to the covariance function defined on and above the sphere. The covariance function, when considered in the plane, is dependent on the Euclidean distance s between two points P and Q :

$$K(P,Q) = K(s) \tag{7-1}$$

with $$s = \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2} \ .$$

Its extension into the halfspace $z \geq 0$ is given by (Moritz, 1976, p.7)

$$K(P,Q) = \int_0^\infty J_o(\eta s) G(\eta) e^{-\eta(z_P + z_Q)} \eta d\eta$$

with z_P, z_Q ... z-coordinates of P, Q ,
$J_o(\eta s)$ ... Bessel function of first kind and zero order
$G(\eta)$ ... Hankel transform of K .

$G(\eta)$ is defined in chapter 2.

In complete analogy to the spherical case , positive definiteness of the function K is equivalent to a positive spectrum $G(\eta)$ for all $\eta \geq 0$. The formulas of the approximating functions (step function, piecewise linear function, cubic spline function) are exactly the same as in the spherical case. The variable t , however, has to be replaced by s (the distance in the plane $z = 0$):

$$f(s)^{(k)} = \sum_{\ell=0}^{L} a_\ell^{(k)} (s - s_k)^\ell \qquad (7\text{-}2)$$

with L = 0 ... step function
L = 1 ... piecewise linear function
L = 3 ... cubic spline function.

(k refers to the subinterval considered.)

We will now investigate, similarly as we did before in the spherical case, how much the spectrum of the approximated covariance function differs from the exact one. Since the Hankel transform $G(\eta)$ is the spectrum of the isotropic covariance function $K(s)$ considered here, we would have to calculate $G(\eta)$. The positivity of $G(\eta)$ is necessary and sufficient for the positive definiteness of $K(s)$. A weaker condition which must be fulfilled by $K(s)$ is the positivity of its Fourier transform

$$F(\omega) = \int_0^\infty K(x)\cos(\omega x)dx \ , \qquad (7\text{-}3)$$

taken along an arbitrary straight line in the horizontal plane $z = 0$. Since all homogeneous and isotropic covariance functions in the plane $z = 0$ are also homogeneous on any straight line in $z = 0$, from the positivity of the Hankel transform $G(\eta) \geq 0$ automatically results the positivity of the Fourier transform $F(\omega)$. Therefore, $F(\omega) \geq 0$ is a necessary condition for a positive definite homogeneous and isotropic function in the plane $z = 0$. The question, whether $F(\omega) \geq 0$ is also a sufficient condition, was put in (Moritz, 1976, p.14). Let us try to give an answer.

7.1 Conditions for positive definiteness of a homogeneous and isotropic covariance function

The geodesist (including the author) is probably more familiar with Legendre polynomials than with Bessel functions, and, since he is usually working on the sphere, also with a discrete spectrum than with a continuous one. Therefore, we first investigate the spherical case and later on the planar one.

7.1.1 Spherical case: the discrete spectrum and the Fourier coefficients

The homogeneous and isotropic covariance function, considered on the unit sphere, has the form

$$K(t) = \sum_{n=2}^{\infty} k_n P_n(t) \qquad (7\text{-}4)$$

with positive coefficients k_n . The spectrum of $K(t)$ results from its projection onto the set of Legendre polynomials (cf.(2-2b)):

$$c_n = \int_{-1}^{1} K(t) P_n(t) dt \qquad (7\text{-}5)$$

with $t := \cos\psi$, ψ... spherical distance.
Due to the mutual orthogonality of Legendre polynomials we obtain

$$c_n = \frac{2}{2n+1} k_n \ , \qquad (7\text{-}6)$$

which is, apart from the constant 2π , the eigenvalue of the covariance function $K(t)$. Since k_n was assumed to be positive, also c_n is positive. The positivity of all c_n is, consequently, a necessary and sufficient condition for a homogeneous and isotropic covariance function on the unit sphere.

The question is, whether also the positivity of all Fourier coefficients

$$f_n = \int_{0}^{\pi} K(\psi) \cos(n\psi) d\psi \qquad (7\text{-}7)$$

is a necessary <u>and</u> sufficient condition for a homogeneous and isotropic positive definite function. (The projection of $K(\psi)$ onto $\sin(n\psi)$ is zero because of the symmetry of $K(\psi)$ with respect to the origin $\psi = 0$.) Taking into account the fact that

$$\cos(n\psi) = \cos(n \cdot \arccos t) = T_n(t) \qquad (7\text{-}8)$$

is the Chebyshev polynomial of first kind and order n , and

observing that

$$d\psi = - \frac{dt}{\sqrt{1-t^2}} ,$$

the Fourier coefficient f_n can be written in the following form:

$$f_n = \int_{-1}^{1} \frac{K(t)T_n(t)}{\sqrt{1-t^2}} \, dt \qquad (7\text{-}9)$$

Since the inner product of Chebyshev polynomials of first kind is defined by

$$\langle T_n, T_m \rangle = \int_{-1}^{1} \frac{T_n(t)T_m(t)}{\sqrt{1-t^2}} \, dt$$

with the non-negative weight function $(1-t)^{-1/2}$, the right hand side of equation (7-9) is exactly the projection of $K(t)$ onto the Chebyshev polynomial $T_n(t)$. Formulas (7-6) and (7-9) are indentical; therefore, the Chebyshev coefficient f_n from (7-9) is identical with the Fourier coefficient f_n from (7-7). Explicitely written, the Fourier (or Chebyshev) coefficients of the covariance function are given by

$$f_n = \int_{-1}^{1} \frac{\left[\sum_{\ell=2}^{\infty} k_\ell P_\ell(t)\right] T_n(t)}{\sqrt{1-t^2}} \, dt \; . \qquad (7\text{-}9)'$$

This equation is a link and allows an insight into the relation between the spectrum c_n of the covariance function, given by eq. (7-5) and its Fourier coefficients. All we have to do is to find a transformation between Legendre and Chebyshev polynomials.

This transformation can easily be found by expressing the Legendre polynomials in terms of powers of its argument t, and then, by expressing the powers of t in terms of Chebyshev polynomials (Davis, 1975, pp.369-372):

$$
\begin{aligned}
P_o &= T_o \\
P_1 &= T_1 \\
P_2 &= \frac{1}{4}(T_o + 3T_2) \\
P_3 &= \frac{1}{8}(3T_1 + 5T_3) \\
P_4 &= \frac{1}{64}(9T_o + 20T_2 + 35T_4) \\
P_5 &= \frac{1}{128}(30T_1 + 35T_3 + 63T_5)
\end{aligned}
\qquad (7\text{-}10)
$$

(The clever reader will find that equation (1-58') in (Heiskanen and Moritz, 1967) is exactly this transformation.) Taking these relations into account, we can express the covariance function $K(t)$ in terms of Chebyshev polynomials

$$
K(t) = \sum_{n=0}^{\infty} d_n T_n(t) \qquad (7\text{-}11)
$$

with coefficients $\{d_n\}$ which can be related to the spectrum $\{c_n\}$. Taking into account the orthogonality of the Chebyshev polynomials with

$$
\langle T_n, T_m \rangle = \begin{cases} 2\pi & \text{for } n = m = 0 \\ \pi & \text{for } n = m \neq 0 \\ 0 & \text{else} \end{cases}, \qquad (7\text{-}12)
$$

we can relate the coefficients d_n to the Fourier coefficients f_n and obtain, after substitution of (7-11) into (7-9)

$$f_n = \sum_{\ell=0}^{\infty} d_\ell \langle T_\ell, T_n \rangle \qquad (7\text{-}13)$$

$$= \begin{cases} \pi d_n & \text{for} \quad n \neq 0 \\ 2\pi d_o & \text{for} \quad n = 0 \ . \end{cases}$$

Therefore, the coefficients d_n in (7-11) are, apart from the constant factor π, identical with the Fourier coefficients f_n.

Comparing (7-11) with (7-4)

$$K(t) = \sum_{n=0}^{\infty} d_n T_n(t) = \sum_{n=2}^{\infty} k_n P_n(t) \ , \qquad (7\text{-}14)$$

we find with the help of the transformation (7-10)

$$P_n(t) = \sum_{m=0}^{n} a_{mn} T_m(t) \qquad (7\text{-}10)'$$

the desired relation between k_n (c_n, respectively) and d_n (f_n, resp.):

$$d_n = \sum_{m=n}^{\infty} a_{nm} k_m \qquad (7\text{-}15)$$

or in matrix notation

$$d = Ak \ ,$$

$$\begin{bmatrix} d_o \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{1}{4} & 0 & \frac{9}{64} & \cdots \\ 0 & 1 & 0 & \frac{3}{8} & 0 & \cdots \\ 0 & 0 & \frac{3}{4} & 0 & \frac{5}{16} & \cdots \\ 0 & 0 & 0 & \frac{5}{8} & 0 & \cdots \\ 0 & 0 & 0 & 0 & \frac{35}{64} & \cdots \\ \cdot & & & & \cdot & \\ \cdot & & & & \cdot & \\ \cdot & & & & \cdot & \end{bmatrix} \begin{bmatrix} k_o \\ k_1 \\ k_2 \\ k_3 \\ k_4 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \qquad (7\text{-}15)'$$

We stated at the beginning that the positivity of all k_n , $n = 2,3,\ldots$ guarantees the positive definiteness of the covariance function given by (7-4). Since all coefficients of A are non-negative and by (7-6) naturally remain non-negative when we use the spectrum $\{c_n\}$ instead of $\{k_n\}$, all Fourier coefficients are positive when all c_n are positive. Therefore, the positivity of the Fourier coefficients is a necessary condition for the positive definiteness of a homogeneous and isotropic function on the sphere.

In order to find an answer whether this condition is also sufficient, we need the inverse of the matrix A . It is very simple to find the inverse: recall that A is the matrix transforming Chebyshev polynomials in Legendre polynomials; consequently, A^{-1} is the matrix transforming Legendre polynomials in Chebyshev polynomials

$$T_n(t) = \sum_{m=0}^{n} a_{mn}^{-1} P_m(t) \ , \qquad (7\text{-}16)$$

explicitely written

$$
\begin{aligned}
T_o &= P_o \\
T_1 &= P_1 \\
T_2 &= \tfrac{1}{3}(-P_o + 4P_2) \\
T_3 &= \tfrac{1}{5}(-3P_1 + 8P_3) \\
T_4 &= \tfrac{1}{105}(-7P_o - 80P_2 + 192P_4) \\
T_5 &= \tfrac{1}{63}(-9P_1 - 56P_3 + 128P_5)
\end{aligned}
\qquad (7\text{-}16)'
$$

Substituting these relations into (7-14) we obtain in analogy to (7-15)

$$
k_n = \sum_{m=n}^{\infty} a^{-1}_{nm} d_m \qquad (7\text{-}17)
$$

$$
k = A^{-1} d \ ,
$$

$$
\begin{bmatrix} k_o \\ k_1 \\ k_2 \\ k_3 \\ k_4 \\ \vdots \end{bmatrix}
\quad
\begin{bmatrix}
1 & 0 & -\frac{1}{3} & 0 & -\frac{1}{15} & \cdots \\
0 & 1 & 0 & -\frac{3}{5} & 0 & \cdots \\
0 & 0 & \frac{4}{3} & 0 & -\frac{16}{21} & \cdots \\
0 & 0 & 0 & \frac{8}{5} & 0 & \cdots \\
0 & 0 & 0 & 0 & \frac{64}{35} & \cdots \\
\vdots & & & & \vdots &
\end{bmatrix}
\quad
\begin{bmatrix} d_o \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ \vdots \end{bmatrix}
\qquad (7\text{-}17)'
$$

The two matrices A and A^{-1} are interesting because of some similarities:

a) both have upper triangular form, therefore, the main diagonal elements are inverse to each other.

b) their sum of column elements is constant and equal to 1; therefore, the following relation holds:

$$\sum k_n = \sum d_n \ ,$$

c) the elements have alternating the value zero and non-zero; consequently, the even k_n depend only on even d_n, odd k_n depend only on odd d_n, and vice versa.

However, apart from these similarities, there is an essential difference between A and A^{-1} : all off-diagonal elements of A^{-1} are zero or negative. Therefore, the positivity of all d_n does, in general, not guarantee the positivity of all k_n and we can state the following:

The positivity of the Fourier coefficients of a homogeneous and isotropic function on the sphere, taken along an arbitrary great circle, is a neccessary but not a sufficient condition for the positive definiteness of the function considered.

According to (7-17)' the set $\{d_n\}$ has, apart from its positivity, to fulfil the following condition in order to guarantee the positive definiteness of the homogeneous and isotropic function on the sphere:

$$d_n > \sum_{k=1}^{\infty} \frac{a^{-1}_{n+2k}}{a^{-1}_{nn}} \, d_{n+2k} \ . \qquad (7\text{-}18)$$

7.1.2 Planar case: the continuous spectrum and the Fourier transform

The fact that a positive Hankel transform causes a positive Fourier transform was already stated at the beginning of this chapter and proved for the spherical case by (7-14)'. It can naturally be expected that similar relations also hold for the planar case; therefore, we will not give a detailed derivation here. However, the inverse question, whether a positive Fourier transform guarantees a positive Hankel transform seems to be of some interest.

We are always concerned with symmetric functions and, therefore, the Fourier transformation reduces from

$$F(\omega) = \int_{-\infty}^{\infty} K(s) e^{-i\omega s} ds$$

to

$$F(\omega) = \int_{-\infty}^{\infty} K(s) \cos(\omega s) ds \ , \qquad (7\text{-}19)$$

with $K(s)$ ... function to be transformed
s ... distance
ω ... frequency.

Because of the symmetry, equation (7-19) can be further simplified to

$$F(\omega) = 2 \int_{0}^{\infty} K(s) \cos(\omega s) ds. \qquad (7\text{-}19)'$$

In the sequel we shall use the first derivative of the Fourier transform which is simply given by

$$F'(\omega) = \frac{dF(\omega)}{d\omega} = -2 \int_{0}^{\infty} K(s) \sin(\omega s) s ds. \qquad (7\text{-}20)$$

In order to obtain an integral representation of the Bessel function of first kind and zero order $J_0(x)$ in terms of a trigonometric

function we have to take advantage to Bessel functions of third kind and zero order which are also referred to as Hankel functions (Lebedev, 1965, p. 107 ff)

$$H_0^{(1)}(x) + H_0^{(2)}(x) = 2J_0(x). \tag{7-21}$$

The corresponding integral representations are given by

$$H_0^{(1)}(x) = \frac{1}{i\pi} \int_{-\infty}^{\infty} e^{ix\cosh\alpha} d\alpha \tag{7-22a}$$

$$H_0^{(2)}(x) = -\frac{1}{i\pi} \int_{-\infty}^{\infty} e^{-ix\cosh\alpha} d\alpha \tag{7-22b}$$

Observing (7-21) we can with the help of (7-22a,b) derive the integral representation of the Bessel function of first kind and zero order

$$J_0(x) = \frac{1}{2i\pi} \int_{-\infty}^{\infty} (e^{ix\cosh\alpha} - e^{-ix\cosh\alpha})\, d\alpha$$

Using the identity

$$\sin(y) = \frac{1}{2i} (e^{iy} - e^{-iy}),$$

the expression above can be simplified to

$$J_0(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \sin(x\cosh\alpha)\, d\alpha .$$

Since $\cosh\alpha$ is symmetric with respect to $\alpha = 0$ we obtain

$$J_0(x) = \frac{2}{\pi} \int_{0}^{\infty} \sin(x\cosh\alpha)\, d\alpha$$

This equation expresses the relation between the Bessel function of first kind and zero order and the trigonometric function sin (·) in full analogy with the spherical case (cf. equation (7-10). Consequently, it provides a link between the Hankel- and the Fourier transform.

Recall the definition of the Hankel transform (2-5)

$$G(\omega) = \int_0^\infty J_0(\omega s)\, K(s)\, s\, ds$$

and substitute (7-23), then we obtain

$$G(\omega) = -\frac{1}{\pi} \int_0^\infty \left[-2 \int_0^\infty K(s) \sin(\omega s \cosh\alpha) ds \right] d\alpha \,.$$

The expression in brackets, however, is according to (7-20) exactly

$$F'(\omega \cosh \alpha),$$

and, consequently, the relation between the Hankel transform $G(\omega)$ and the Fourier transform $F(\omega)$ is given by

$$G(\omega) = -\frac{1}{\pi} \int_0^\infty F'(\omega \cosh \alpha)\, d\alpha \,. \qquad (7\text{-}24)$$

Since the positivity of the Hankel transform guarantees the positive definiteness of a homogeneous and isotropic function in the plane z=0, the Fourier transform has, apart from its positivity, to fulfil

$$-\int_0^\infty F'(\omega \cosh\alpha)\, d\alpha > 0.$$

7.2 Fourier transforms of covariance function approximations in the plane

For three reasons we leave now the Hankel transformation and work with the Fourier transformation:

a) We will investigate the problem of covariance function approximations by finite element models; therefore, we are primarily interested if these approximating functions fulfil the necessary condition for positive definiteness which is defined via the Fourier transform and, if not, what the degree of non-positivity is.

b) For numerical studies we choose a very simple model, the Gaussian covariance function. It can be shown (Moritz, 1976, p.19) that, apart from a constant factor, the Fourier and the Hankel transform coincide for this model. Therefore, it can be expected that the Fourier transform of the approximating function does not differ much from the corresponding Hankel transform.

c) The calculation of the Fourier transform is much easier than that of the Hankel transform: in the latter the integral kernel consists basically of the Bessel function of first kind and zero order which can, for arbitrary argument, only be calculated by numerical integration; in the former case, the integral kernel is a simple trigonometric function.

We will, as in the spherical case, investigate the step function, piecewise linear function and the cubic spline function as approximating elements. Because of the symmetry of the functions with respect to $s = 0$, also the Fourier transform is symmetric with respect to the Frequency $\omega = 0$, and moreover, the imaginary part vanishes identically. Consequently, the Fourier transform which will be denoted by "spectrum" in the sequel is given by

$$F(\omega) = 2\int_0^\infty K(s)\cos(\omega s)ds \ .$$

The different approximating functions are put in a closed form by (7-2)

$$f^{(j)}(s) = \sum_{\ell=0}^{L} a_\ell^{(j)}(s-s_j)^\ell \ ,$$

where L assumes its maximum for the cubic spline $(L = 3)$. Therefore, we will derive the spectrum for this function; the spectra of the other (inferior) functions can then be obtained after specialisation.

Theoretically, we should calculate the integral over the whole domain $-\infty < s < \infty$; however, since practically all covariance functions tend to zero quickly, we can limit ourselves to a finite interval. This interval will be subdivided into J subintervals bounded by the mesh points s_j , s_{j+1} . The interpolating function defined on this subinterval $[s_j , s_{j+1}]$ will be denoted by $f^{(j)}(s)$; its equation is given above.

We calculate, therefore, the Fourier transform of the interpolating (approximating) function $f(s)$ consisting of functions $f^{(j)}(s)$

$$F(\omega) = 2 \sum_{j=1}^{J} \int_{s_j}^{s_{j+1}} f^{(j)}(s) \cos(\omega s) ds \ . \qquad (7\text{-}26)$$

Substituting for $f^{(j)}(s)$ we obtain

$$F(\omega) = 2 \sum_{j=1}^{J} \sum_{\ell=0}^{L} a_\ell^{(j)} C_\ell^{(j)}(\omega) \qquad (7\text{-}27a)$$

$$C_\ell^{(j)}(\omega) = \int_{s_j}^{s_{j+1}} (s-s_j)^\ell \cos(\omega s) ds \ . \qquad (7\text{-}27b)$$

For reasons of a compact notation we also define the functions

$$S_\ell^{(j)}(\omega) = \int_{s_j}^{s_{j+1}} (s-s_j)^\ell \sin(\omega s) ds \ . \qquad (7\text{-}27c)$$

Performing the integrations in (7-27b,c) and putting

$$\Delta_j := s_{j+1} - s_j$$

we arrive at the following expressions:

$$
\begin{aligned}
C_o^{(j)}(\omega) &= \frac{1}{\omega}\left[\sin(\omega s_{j+1}) - \sin(\omega s_j)\right] \\
S_o^{(j)}(\omega) &= -\frac{1}{\omega}\left[\cos(\omega s_{j+1}) - \cos(\omega s_j)\right] \\
C_1^{(j)}(\omega) &= -\frac{1}{\omega}\left[S_o^{(j)}(\omega) - \Delta_j \sin(\omega s_{j+1})\right] \\
S_1^{(j)}(\omega) &= \frac{1}{\omega}\left[C_o^{(j)}(\omega) - \Delta_j \cos(\omega s_{j+1})\right] \\
C_2^{(j)}(\omega) &= -\frac{1}{\omega}\left[2S_1^{(j)}(\omega) - \Delta_j^2 \sin(\omega s_{j+1})\right] \\
S_2^{(j)}(\omega) &= \frac{1}{\omega}\left[2C_1^{(j)}(\omega) - \Delta_j^2 \cos(\omega s_{j+1})\right] \\
C_3^{(j)}(\omega) &= -\frac{1}{\omega}\left[3S_2^{(j)}(\omega) - \Delta_j^3 \sin(\omega s_{j+1})\right] \\
S_3^{(j)}(\omega) &= \frac{1}{\omega}\left[3C_2^{(j)}(\omega) - \Delta_j^3 \cos(\omega s_{j+1})\right] .
\end{aligned}
\tag{7-28}
$$

With the help of these relations we are able to calculate the Fourier spectrum for all functions chosen here. For our numerical example we have used the Gaussian covariance function of gravity anomalies defined by

$$
C(s) = C_o e^{-\ln 2 \left(\frac{s}{h}\right)^2} \tag{7-29}
$$

with the following parameters:

$C_o = 1600\ \text{mgal}^2$ ... variance of gravity anomalies
$h = 60\ \text{km}$ ... correlation length.

(the variance G_o of the horizontal derivatives of the gravity anomalies Δg assumes the value $G_o \doteq 60\ E^2$.)

The Fourier transform of this model is given by (Moritz, 1976, p.19)

$$F(\omega) = C_o h \sqrt{\frac{\pi}{\ln 2}}\, e^{-\frac{(\omega h)^2}{4 \ln 2}} \qquad (7\text{-}30)$$

For comparison purposes we give also the Hankel transform which differs from (7-30) only by a scale factor:

$$G(\omega) = C_o \frac{h^2}{2 \ln 2}\, e^{-\frac{(\omega h)^2}{4 \ln 2}} \quad .$$

Similarly as in chapter 6 we define the ralative perturbation of the spectrum by

$$q(\omega) := \left| 1 - \frac{\tilde{F}(\omega)}{F(\omega)} \right|$$

with $F(\omega)$ ... exact spectrum (eq. 7-30)
$\tilde{F}(\omega)$ ... spectrum of the approximating function.

For the three models discussed here, $q(\omega)$ is shown in figure 7.1. According to the definition of $\omega = 2\pi/T$ (T ... period), $\omega = 1$ corresponds to a wavelength of 2π km when the distance s is counted in kilometers. Therefore, $\omega = 0.1$ corresponds to $2\pi \cdot 10$ km which is approximately the correlation length chosen here. The subinterval length Δ was set constant and equal to 30 km (1/2 of the correlation length h).

From figure 7.1 we can draw the following conclusions: For the spline approximation the perturbation of the spectrum is practically zero up to $\omega = 0.05$ which corresponds to a wavelength of h/2 ; until $\omega = 0.1$ it increases modestly and

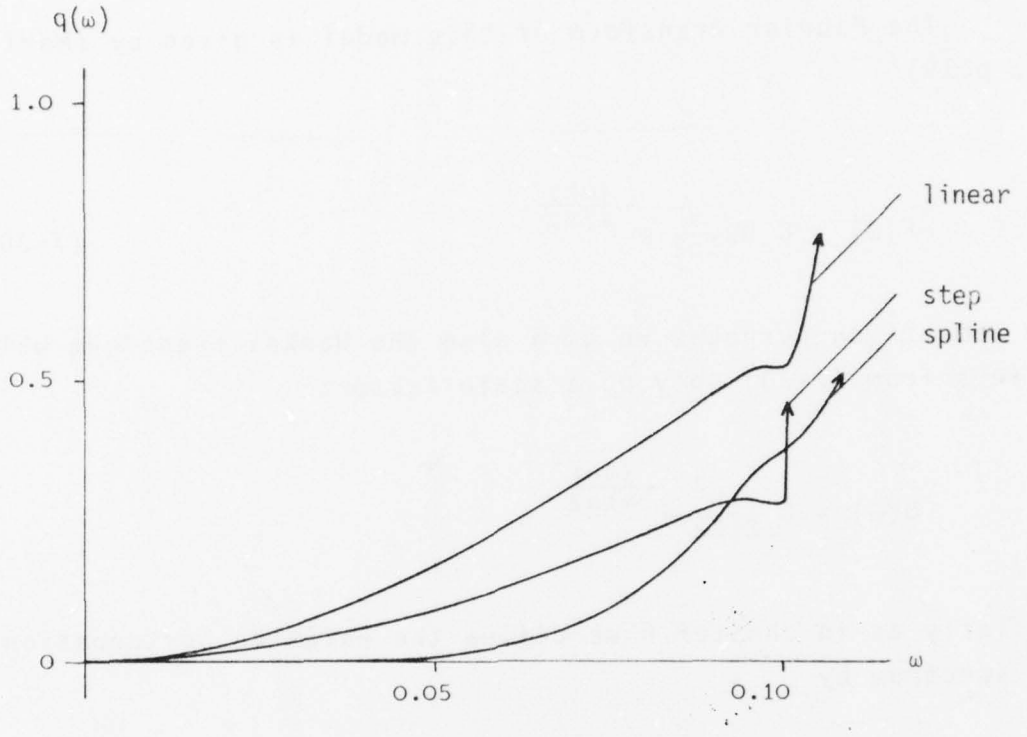


Fig. 7.1 Relative perturbation of the spectrum (planar case)

from 0.1 on it is falsified too much. The vanishing perturbation for small ω expresses the fact that the low frequency part of the spectrum is very well represented by the spline. The situation is different for the piecewise linear function and the step function where the perturbation is already considerable in the very low frequency part which is not sufficiently well represented. Similar as for the spline, the perturbation becomes very big from $\omega = 0.1$ on. From the first view the reader will probably conclude that the author has made a mistake by changing the curves for the piecewise linear function and the step function. However, this is not the case. The reason that the low frequency part of the spectrum is worse represented by the piecewise linear function lies in the fact that for small s (the most important part of the covariance function) the interpolating linear function remains below the exact one and, therefore, causes an under-represented low frequency part of the spectrum. In order to show

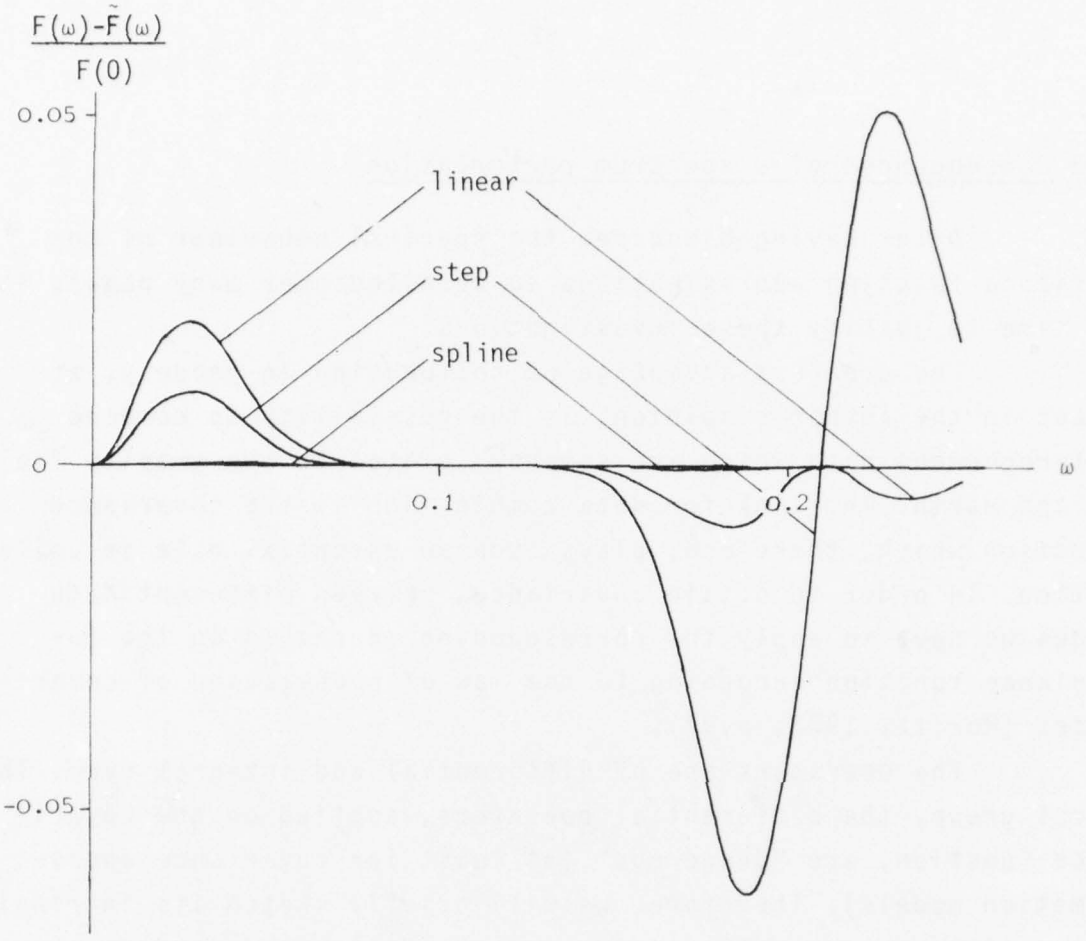


Fig. 7.2 Perturbation of the spectrum normalized by F(0)

the difference between the exact spectrum and the spectrum of the approximating function we have drawn figure 7.2. It shows the deviation of the two spectra normalized by $F(0)$.

Whereas the low frequency part of the spectrum is better represented by the step function than by the piecewise linear function, it is much worse in the higher frequency part (this strong perturbation comes mainly from the discontinuity of the approximating function). For the spline the relative perturbation remains below $5 \cdot 10^{-4}$.

7.3 Consequences of a spectrum perturbation

After having discussed the spectral behaviour of covariance function approximations so detailed over many pages, it is time to justify these investigations.

The greatest advantage of collocation in geodesy, at least in the author's opinion, is the possibility to combine heterogeneous data which are somehow related to the gravity field of the earth. The link for data combination is the covariance function which, therefore, plays such an essential role in collocation. In order to obtain covariances between different data types we have to apply the corresponding operators on the covariance function according to the law of propagation of covariances (Moritz, 1972, p.97).

The operators are of differential and integral type. The first group, the differential operators, applied on the covariance function, are "dangerous" (at least for covariance approximation models). Therefore, we will briefly sketch its intrinsic properties. We are only considering the distance dependent covariance function here.

a) Differentiation

Let us again consider the covariance function of the potential in its basic form on the unit sphere

$$K(t) = \sum_n k_n P_n(t)$$

and let us denote by $\tilde{K}(t)$ its approximation by some model with corresponding coefficients $\tilde{k}_n$. Take, for example, the cross covariance between the geoidal height and the deviation of the vertical. Then we have, besides other trivial manipulations, to differentiate the covariance function with respect to the independent variable t

$$K'(t) = \sum_n k_n P_n'(t) \tag{7-31}$$

with $$P_n'(t) = -\frac{n}{1-t^2}\left[tP_n(t) - P_{n-1}(t)\right] . \tag{7-32}$$

The analogue expression for the approximation is

$$\tilde{K}'(t) = \sum_n \tilde{k}_n P_n'(t) . \tag{7-33}$$

Equation (7-32) tells us the principal action of differentiation: it acts as an amplifier. The higher the degree n, the higher will be the amplification because, basically, k_n is multiplied by n. Consider, now, the error

$$\Delta'(t) := K'(t) - \tilde{K}'(t) = \sum_n (k_n - \tilde{k}_n) P_n'(t)$$

$$= -\sum_n (k_n - \tilde{k}_n)\frac{n}{1-t^2}\left[tP_n(t) - P_{n-1}(t)\right] . \tag{7-34}$$

It is obvious that the error $\Delta'(t)$ is very sensitive to the error $k_n - \tilde{k}_n$ which is related to the perturbation of the spectrum by (2-3)'.

Similar as in the spherical case is the situation in the plane. Let the covariance function be defined by $C(s)$ via the Hankel transform of its spectrum (the Hankel transform is involutoric for quadratically integrable functions in the plane)

$$C(s) = \int_0^\infty J_o(\omega s) G(\omega) \omega d\omega . \tag{7-35}$$

Its differentiation with respect to s is given by

$$C'(s) = \int_0^\infty \frac{\partial J_o(\omega s)}{\partial s} G(\omega)\omega d\omega \ ,$$

which, with the help of (Papoulis, 1968, p.168)

$$\frac{\partial J_o(\omega s)}{\partial s} = - \omega J_1(\omega s) \qquad (7\text{-}36)$$

($J_1(x)$ is the Bessel function of first kind and first order) assumes the form

$$C'(s) = - \int_0^\infty J_1(\omega s) G(\omega) \omega^2 d\omega \ . \qquad (7\text{-}37)$$

Since $J_1(x)$ has, basically, the same amplitude behaviour as $J_o(x)$, the term ω^2 (instead of ω in the usual transformation) acts as amplifier and plays, therefore, the same role as n in the spherical case. The error of the differentiated covariance function is, in analogy to (7-34) given by

$$\Delta'(s) = C'(s) - \tilde{C}'(s) = - \int_0^\infty J_1(\omega s) \left[G(\omega) - \tilde{G}(\omega)\right] \omega^2 d\omega \qquad (7\text{-}38)$$

with $\tilde{C}(s)$ ... approximated covariance function,
$\tilde{G}(\omega)$ ... its corresponding Hankel transform.

In view of these facts we conclude that it is extremely important to know, up to which frequency the spectrum of the approximating function is well behaved and at which frequency it starts to become falsified. Since the perturbation of the spectrum depends on the smoothness of the interpolating element used and on the

length of the subinterval chosen, which was shown in the last chapters, we can select the proper function together with a certain grid spacing as soon as we know which collocation problem we are going to solve. Especially when differential operators are involved, the study of the perturbation of the spectrum is of considerable importance.

b) Integration

More pleasant,as far as errors are concerned, are problems which involve integrations of the covariance function. In contrast to the differentiation, the integration has a smoothing effect and, consequently, we are much more free regarding the choice of the interpolating element and the length of the subinterval.

Take, for example, the integral of the covariance function over the interval $[t_o, t_1]$,

$$\int_{t_o}^{t_1} K(t)dt = \sum_n k_n \int_{t_o}^{t_1} P_n(t)dt ,$$

then the corresponding approximation error is given by

$$\int_{t_o}^{t_1} \Delta(t)dt = \sum_n \frac{k_n - \tilde{k}_n}{2n+1} \left[P_{n+1}(t_1) - P_{n-1}(t_1) + P_{n-1}(t_o) - P_{n+1}(t_o) \right] . \qquad (7\text{-}39)$$

Analogous formulas can be obtained for the planar case

$$\int_{s_o}^{s_1} \Delta C(s)ds = \int_{\omega=0}^{\infty} \int_{s=s_o}^{s_1} J_o(\omega s)\Delta G(\omega)\omega d\omega ds ,$$

which, however, lead to rather complicated expressions involving Struve functions (Sünkel, 1977, p.133). For the Fourier transform it is much easier to give an error estimate

$$\int_{s_o}^{s_1} \Delta C(s)ds := 2 \int_{\omega=0}^{\infty} \int_{s=s_o}^{s_1} \cos(\omega s)\Delta F(\omega)ds d\omega ,$$

which can be simplified by considering (7-28) :

$$\int_{s_o}^{s_1} \Delta C(s)ds = 2\int_{0}^{\infty} \frac{1}{\omega}\left[\sin(\omega s_1) - \sin(\omega s_o)\right]\Delta F(\omega)d\omega . \quad (7\text{-}40)$$

The smoothing properties of the integral operator are evident from the fact, that in (7-39) and in (7-40) the frequency occurs in the denominator and, therefore, damps errors in the high frequency part of the spectrum. Consequently, for collocation problems involving integrations of the covariance function, even such approximating functions can be used which give rise to relatively big perturbations in the high frequency part of the spectrum, like the step function (cf. fig.7.2). The treatment of operators like differential and integral operators in the spectral domain is, therefore, very well suited to give an idea of its intrinsic properties. After having knowledge about the problem to be solved, we can directly select the proper approximation element together with the length of the subinterval. This is an important completion of the error estimates given in chapters 3, 4 and 5.

8. Errors of prediction caused by approximations

In the preceding chapters we have tried to give rules of thumb for the errors which have to be expected for different covariances, being elements of the covariance matrix. The underlying principle was bilateral: on the one hand we gave maximum errors derived directly from the approximating function chosen, on the other hand the estimations were obtained via the study of

the perturbations of the spectrum. Both results supplement each other. These investigations were absolutely necessary because they provided us the information about the errors of the elements of the covariance matrix and the signal-measurement covariance vector.

The approximation, so to say, perturbs the elements of the linear transformation between measurement x and predicted signal s_P. The two formulas we are primarily interested are (Moritz, 1970, p.7)

$$s_P = C_P^T C^{-1} x \ , \tag{8-1}$$

$$m_P^2 = C_{PP} - C_P^T C^{-1} C_P \tag{8-2}$$

with
- x ... vector of measurements
- C ... covariance matrix of measurements
- s_P ... predicted signal
- C_P ... covariance vector between signal s_P and measurement x
- m_P^2 ... mean square error of the predicted signal s_P
- C_{PP} .. variance of the predicted signal s_P .

(We do neither consider a systematic part nor noise in the measurements.)

By the approximation of the basic covariance function, in general, all covariances are disturbed to some extent. In the following we make some attempts to estimate the consequences of the approximation for the predicted signal and its mean square error.

8.1 Error norm estimates

Let us consider covariances being elements of $\tilde{C}$ and $\tilde{C}_P$ which differ slightly from its exact analogues C and C_P :

$$C = \tilde{C} + \delta C \ , \qquad (8\text{-}3)$$

$$C_P = \tilde{C}_P + \delta C_P \ .$$

Then it can be expected that also the signal $\tilde{s}_P$ resulting from the prediction process (8-1) differs only slightly from s_P :

$$s_P = \tilde{s}_P + \delta s_P \ . \qquad (8\text{-}4)$$

With (8-1) - (8-4) the signal approximation error δs_P can be obtained by

$$\delta s_P = (C_P^T C^{-1} - \tilde{C}_P^T \tilde{C}^{-1}) x \ . \qquad (8\text{-}5)$$

The last expression within the brackets can be split according to (8-3) into

$$\tilde{C}_P^T \tilde{C}^{-1} = (C_P^T - \delta C_P^T)(C - \delta C)^{-1} \ . \qquad (8\text{-}6)$$

If the matrix C is reasonably stable, we can expect that the inverse $\tilde{C}^{-1}$ is close to the exact one C^{-1} :

$$(C - \delta C)^{-1} = C^{-1} + \delta K \qquad (8\text{-}7)$$

with δK small relativ to C^{-1} . Consequently, the following relation holds with sufficient accuracy:

$$\begin{aligned} (C - \delta C)(C^{-1} + \delta K) &= I \\ &= CC^{-1} - \delta C C^{-1} + C\delta K - \delta C \delta K \\ &\doteq I - \delta C C^{-1} + C\delta K \ , \end{aligned}$$

which gives

$$\delta K = C^{-1}\delta C C^{-1} \quad , \tag{8-8}$$

so that

$$(C - \delta C)^{-1} \doteq C^{-1}(I + \delta C C^{-1}) \quad , \tag{8-9}$$

with I denoting the unit matrix. Analogously we simplify equation (8-6)

$$\begin{aligned} \tilde{C}_P^T \tilde{C}^{-1} &= (C_P^T - \delta C_P^T) C^{-1}(I + \delta C C^{-1}) \\ &\doteq C_P^T C^{-1} - \delta C_P^T C^{-1} + C_P^T C^{-1}\delta C C^{-1} \quad , \end{aligned}$$

so that the following relation holds for the signal approximation error:

$$\delta s_P = (\delta C_P^T - C_P^T C^{-1}\delta C)C^{-1}x \quad . \tag{8-5)'}$$

In a similar way also the approximation error of the mean square error can be found. Let us denote this error by

$$\delta m_P^2 := m_P^2 - \tilde{m}_P^2 \tag{8-10}$$

and the approximation error of the signal variance by

$$\delta C_{PP} := C_{PP} - \tilde{C}_{PP} \quad .$$

If we neglect, as above, products of small terms, we obtain

$$\delta m_P^2 = \delta C_{PP} - \delta C_P^T C^{-1} C_P + C_P^T C^{-1}\delta C C^{-1} C_P - C_P^T C^{-1}\delta C_P \quad .$$

Since a real number does not change under transposition, the following relation holds:

$$\delta C_P^T C^{-1} C_P = C_P^T C^{-1} \delta C_P \ ,$$

so that the above expression for δm_P^2 can be further simplified:

$$\delta m_P^2 = \delta C_{PP} - (2\delta C_P^T - C_P^T C^{-1} \delta C) C^{-1} C_P \ . \qquad (8\text{-}10)'$$

From the equations (8-5)' and (8-10)' we can already conclude that it is impossible to give an idea about the approximation errors of the signal and the mean square error as long as information about C^{-1} is not available. C, however, depends on the data configuration, and so does C^{-1}. Therefore, it is not easy to give error estimates without having information about the intrinsic properties of the covariance matrix.

Introducing some kind of norm denoted by $||\cdot||$ and recalling the relations

$$||AB|| \leq ||A|| \ ||B||$$

we can give upper bounds for the errors δs_P and δm_P^2 :

$$|\delta s_P| \leq (||\delta C_P^T|| + ||C_P^T|| \ ||C^{-1}|| \ ||\delta C||) \ ||C^{-1}|| \ ||x|| \ , \qquad (8\text{-}11)$$

$$|\delta m_P^2| \leq ||\delta C_{PP}|| + (2||\delta C_P^T|| + ||C_P^T|| \ ||C^{-1}|| \ ||\delta C||) \cdot \cdot ||C^{-1}|| \ ||C_P|| \ . \qquad (8\text{-}12)$$

The following example may give an indication how pessimistic these error bounds are:
Let us consider two points P_1 and P_2 separated by the distance $s = 1$, and measurements x_1 and x_2 at these points. Predict the signal s_P, which is of the same type as x_1 and x_2, at the midpoint of the straight line connecting P_1 and P_2. The covariance function is Gaussian and given by $C(s) = \exp(-s^2)$. The exact covariances have the values

$$C = \begin{bmatrix} 1 & 0.367879 \\ 0.367879 & 1 \end{bmatrix} , \quad C_P^T = (0.778801 , 0.778801) .$$

As approximations we choose

$$\tilde{C} = \begin{bmatrix} 1 & 0.36 \\ 0.36 & 1 \end{bmatrix} , \quad \tilde{C}_P^T = (0.77 , 0.77) ;$$

the data have the values

$$x^T = (1, 2) .$$

Performing the trivial calculations according (8-1) and (8-2) we obtain the following results:

$$s_P = 1.708 \qquad m_P^2 = 0.113$$
$$\tilde{s}_P = 1.699 \qquad \tilde{m}_P^2 = 0.128 ,$$

so that the errors assume the values

$$|\delta s_P| = 0.009 \qquad |\delta m_P^2| = 0.015 .$$

With the max-norm

$$||x|| = \max_i |x_i| , \qquad x \ldots \text{vector},$$

$$||A|| = \max_i \sum_k |a_{ik}| , \qquad A \ldots \text{matrix},$$

which are very easy to calculate, we obtain the estimates

$$|\delta s_P| \leq 0.059 \qquad |\delta m_P^2| \leq 0.034 ,$$

which are obviously very pessimistic compared to the exact differences calculated above. Worse results are obtained with the Euclidean norm

$$||x||^2 = x^T x , \qquad x \ldots \text{vector}$$

$$||A||^2 = \mathrm{tr}(A^T A) , \qquad A \ldots \text{matrix} \quad (\mathrm{tr} \ldots \text{trace}):$$

$$|\delta s_P| \leq 0.132 \qquad |\delta m_P^2| \leq 0.089 .$$

Finally we consider the estimates provided by the spectral norm defined by

$$||x|| = (x^T x)^{1/2} , \qquad x \ldots \text{vector}$$

$$||A|| = \kappa_{max} , \qquad A \ldots \text{matrix}$$

with $\kappa^2 := \lambda_{A^T A}$, λ ... eigenvalue. Since, in our case, all matrices are symmetric, the product

$$A^T A = AA$$

and the eigenvalues

$$\lambda_{A^T A} = \lambda_{AA} = \lambda_A^2 ,$$

so that

$$||A|| = \lambda_{max}$$

holds. For this norm we obtain the estimates

$$|\delta s_P| \leq 0.094 \qquad |\delta m_P^2| \leq 0.067 .$$

As long as matrices are involved, the spectral norm gives the smallest value; if, however, vector norms enter into the estimation, the max-norm defined above is, for vectors, smaller or equal to the Euclidean one. This is the reason why the estimates for the max-norm obtained here are smallest. This, however, cannot be generalized.

At this point it seems to be worth to put the question, which condition has to be fulfilled by the approximated covariances in order to guarantee the following basic condition:

$$0 \leq \tilde{m}_P^2 \leq \tilde{C}_{PP} \qquad (8\text{-}13)$$

The first condition is equivalent to non-negative mean square error, so that no imaginary root mean square error can occur. (It looks trivial, but it is essential; some polynomial approximations of the covariance function do not fulfil this condition.) The second one is also necessary since the error variance cannot be bigger than the variance of the signal. (Adding one single measurement, the mean square error has to decrease.)

Both questions can be answered immediately:

$$0 \leq \tilde{C}_P^T \tilde{C}^{-1} \tilde{C}_P \leq \tilde{C}_{PP} \qquad (8\text{-}13)$$

is a necessary and sufficient condition for (8-13). $\tilde{C}_P^T \tilde{C}^{-1} \tilde{C}_P > 0$ is true if the covariance matrix $\tilde{C}$ is positive definite. (We neglect here the equality sign and do not speak about semi-definiteness.) Positive definiteness, however, is equivalent to positive eigenvalues; therefore, the spectrum enters here again.

8.2 Perturbation of eigenvalues

Let us again consider the covariance matrix C which is disturbed slightly giving C_ε and suppose that C_ε depends on a parameter ε in such a way that it reduces to $\tilde{C}$ for $\varepsilon = 1$ (the approximated covariance matrix in consideration) and to C for $\varepsilon = 0$. We follow here(Friedrichs, 1973, pp.213). Let C_ε be analytic in ε; it then admits a representation

$$C_\varepsilon = C + \varepsilon C_1 + \varepsilon^2 C_2 + \ldots \qquad (8\text{-}14a)$$

with bounded $C_1, \ldots$. In the same way we expand the eigenvalues λ_ε and corresponding eigenvectors y_ε of C_ε:

$$\lambda_\varepsilon = \lambda + \varepsilon \lambda_1 + \varepsilon^2 \lambda_2 + \ldots \qquad (8\text{-}14b)$$

$$y_\varepsilon = y + \varepsilon y_1 + \varepsilon^2 y_2 + \ldots \qquad (8\text{-}14c)$$

($\lambda_\varepsilon = \tilde{\lambda}$ for $\varepsilon = 1$, $\lambda_\varepsilon = \lambda$ for $\varepsilon = 0$; $y_\varepsilon = \tilde{y}$ for $\varepsilon = 1$, $y_\varepsilon = y$ for $\varepsilon = 0$.)

The relation between C_ε, λ_ε and y_ε is called the eigenvalue problem

$$(C_\varepsilon - \lambda_\varepsilon)y_\varepsilon = 0 \ . \tag{8-15}$$

Substituting (8-14a-c) into (8-15) we obtain

$$(C + \varepsilon C_1 + \varepsilon^2 C_2 + \ldots)(y + \varepsilon y_1 + \varepsilon^2 y_2 + \ldots)$$

$$= (\lambda + \varepsilon\lambda_1 + \varepsilon^2\lambda_2 + \ldots)(y + \varepsilon y_1 + \varepsilon^2 y_2 + \ldots) \ .$$

We order with respect to powers of ε resulting in a sequence of equations:

$$\begin{aligned} (C - \lambda)y &= 0 \\ (C - \lambda)y_1 &= -(C_1 - \lambda_1)y \\ (C - \lambda)y_2 &= -(C_1 - \lambda_1)y_1 - (C_2 - \lambda_2)y \\ &\vdots \end{aligned} \tag{8-16}$$

Here we restrict ourselves to expressions linear in ε . Since C and its disturbance is symmetric, we note that

$$y^T[(C - \lambda)z] = 0 \tag{8-17}$$

holds for every vector z . This follows from

$$y^T[(C - \lambda I)z] = z^T[(C - \lambda I)y] = z^T 0 = 0$$

because y is eigenvector of C . Therefore, we premultiply the second equation of (8-16) by y^T and obtain

$$y^T[(C - \lambda I)y_1] = -\, y^T[(C_1 - \lambda_1 I)y] \ ,$$

which gives zero because of (8-17):

$$y^T[(C_1 - \lambda_1 I)y] = 0 \ .$$

This equation can be split into

$$y^T C_1 y = \lambda_1 y^T y \ ,$$

and we can express λ_1 in terms of the eigenvector y and the disturbation matrix C_1 ,

$$\lambda_1 = \frac{y^T C_1 y}{y^T y} \ . \qquad (8\text{-}18)$$

If y is a normalized eigenvector, (8-18) can be simplified to

$$\lambda_1 = y^T C_1 y \ .$$

Translated into our problem (C_1 corresponds to $-\delta C$, λ_1 corresponds to $-\delta\lambda$) it has the form

$$\delta\lambda = y^T \delta C y \qquad (8\text{-}19)$$

with (ε is put equal to 1 now)

$$\tilde{\lambda} = \lambda - \delta\lambda \ , \quad \tilde{C} = C - \delta C \ . \qquad (8\text{-}20)$$

It is fairly easy to give an upper bound for the eigenvalue perturbation $\delta\lambda$ when the spectral norm is used. Since the eigenvector y is assumed to be normalized it has the Euclidean norm (length) equal to 1

$$||y|| = 1$$

and the norm of δC is given by its largest eigenvalue

$$||\delta C|| = \max|\lambda_{\delta C}| \ ,$$

so that

$$|\delta\lambda| \leq \max|\lambda_{\delta C}| \ .$$

The norm $|\delta\lambda|$, again, plays a fundamental role, because, according to (8-20), its value relative to λ tells us whether we can be sure that all eigenvalues of $\tilde{C}$ are positive (what is equivalent to positive definiteness) or not. If

$$\max|\lambda_{\delta C}| > \min \lambda_C \ , \tag{8-21}$$

we have to be very careful as far as the positive definiteness of $\tilde{C}$ is concerned (this is an essential condition).

An idea about the variation of λ_C and $\lambda_{\delta C}$ might be provided for some stable problems by the third theorem of Gershgorin (Wilkinson, 1965,p.71). According to this theorem, all eigenvalues λ_i ($i = 1,\ldots, n$) of a ($n \times n$) matrix A are situated within the union of the discs (Fig.8.1)

$$|z - a_{ii}| \leq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \ . \tag{8-22}$$

When A is symmetric, the complex number z reduces to a real one and, consequently, the disc degenerates to an interval on the real line. In view of the Gershgorin theorem the condition for positive definiteness of $\tilde{C}$ can be put in the form

$$\max_i\left|\delta c_{ii} + \sum_{\substack{j=1 \\ j\neq i}}^{n} |\delta c_{ij}|\right| < \min_i\left(c_{ii} - \sum_{\substack{j=1 \\ j\neq i}}^{n} |c_{ij}|\right) \tag{8-23}$$

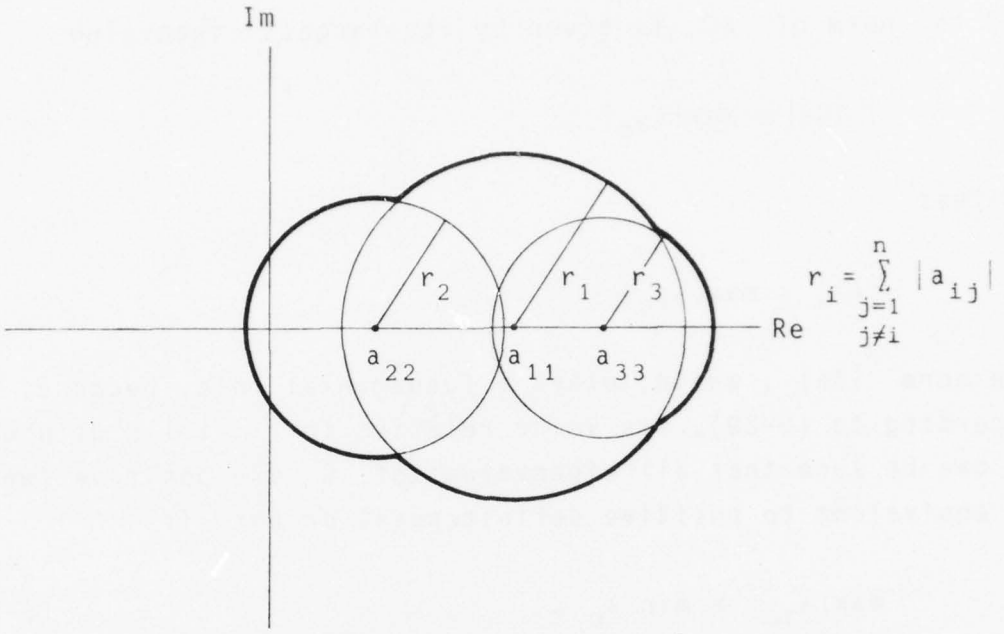


Fig. 8.1 Third theorem of Gershgorin

For the example discussed above this condition is certainly fulfilled:

$$\max_i \left| \delta c_{ii} + \sum_{\substack{j=1 \\ j\neq i}}^{n} |\delta c_{ij}| \right| = 0.007879 ,$$

$$\min_i \left(c_{ii} - \sum_{\substack{j=1 \\ j\neq i}}^{n} |c_{ij}| \right) = 0.632121 .$$

8.3 Stochastic error estimates

Up to now the error estimates were derived from a purely deterministic approximation procedure without any stochastic interpretation. Although in such an approximation problem there is really no stochastic background, it might be sometimes (at least when large problems are treated) useful and appropriate to consider the disturbances of the covariances as a sample of a stochastic process. In this way the elements of δC_P and δC

are considered as random variables of a second order stochastic process uniquely described by the expectation $E\{\cdot\}$ and the variance $E\{\cdot\cdot\}$ *.

Let us recall equation (8-5)':

$$\delta s_P = (\delta C_P^T - C_P^T C^{-1} \delta C) C^{-1} x \ .$$

We change the notation such that

$$\sigma = \delta s_P \ , \quad \theta = \delta C_P \ , \quad a^T = - C_P^T C^{-1} \ , \quad \Theta = \delta C \ , \quad b = C^{-1} x \ ;$$

with these notations the equation above can be written in the form

$$\sigma = (\theta + a^T \Theta) b \qquad (8\text{-}24)$$

where a, b ... deterministic vectors
θ ... stochastic vector
Θ ... stochastic matrix.

For the statistics of θ and Θ we make the following simplifying assumptions:

$$E\{\theta_i\} = 0 \ , \qquad E\{\theta_i \theta_j\} = \alpha^2 \delta_{ij} \ ,$$

$$E\{\Theta_{ij}\} = 0 \ , \qquad E\{\Theta_{ij} \Theta_{k\ell}\} = \alpha^2 \delta_{ik} \delta_{j\ell} \ , \quad ** \qquad (8\text{-}25)$$

$$E\{\theta_i \Theta_{ik}\} = 0 \ .$$

---

* ... E denotes expectation in the probabilistic sense,
** ... δ_{ij} is the Kronecker symbol .

With these assumptions it is easy to show that the expectation of σ vanishes:

$$E\{\sigma\} = E\{\theta_i\}b_i + a_i E\{\Theta_{ij}\}b_j = 0 \ .^{*}$$

For the variance of σ we derive

$$E\{\sigma^2\} = E\{(\theta_i b_i + a_i \Theta_{ij} b_j)(\theta_k b_k + a_k \Theta_{k\ell} b_\ell)\}$$

$$= E\{\theta_i \theta_k\} b_i b_k + 2E\{\Theta_{ij}\theta_k\} a_i b_j b_k$$

$$+ E\{\Theta_{ij}\Theta_{k\ell}\} a_i a_k b_j b_\ell \ ,$$

which, by (8-25), becomes

$$E\{\sigma^2\} = \alpha^2 \delta_{ik} b_i b_k + \alpha^2 \delta_{ik}\delta_{j\ell} a_i a_k b_j b_\ell$$

$$= \alpha^2 (b_i b_i + a_i a_i b_j b_j)$$

or $$E\{\sigma^2\} = \alpha^2 b^T b(1 + a^T a) \ .$$

After back-substitution for a and b we finally get

$$E\{\delta s_P \delta s_P\} = \alpha^2 x^T C^{-1} C^{-1} x(1 + C_P^T C^{-1} C^{-1} C_P) \ . \qquad (8\text{-}26)$$

This is the mean square error of the predicted signal due to the approximation with the assumption made above. Similarly we can derive the mean square approximation error of the mean square error of the prediction. Recall (8-10)'

$$\delta m_P^2 = \delta C_{PP} - (2\delta C_P^T - C_P^T C^{-1} \delta C) C^{-1} C_P \ ,$$

---

* ... We make use of the summation convention.

which, with $\mu := \delta m_P^2$, $\gamma := \delta C_{PP}$ and θ , Θ, a as before, can be written as

$$\mu = \gamma + a^T(2\theta + \Theta a) , \tag{8-27}$$

where γ is a stochastic scalar with

$$\begin{aligned} E\{\gamma\} &= 0 , \qquad E\{\gamma^2\} = \alpha^2 , \\ E\{\gamma\theta_i\} &= 0 , \qquad E\{\gamma\Theta_{ij}\} = 0 . \end{aligned} \tag{8-28}$$

As before it can immediately be verified that $E\{\mu\} = 0$. For the variance we derive

$$\begin{aligned} E\{\mu^2\} &= E\{(\gamma + 2a_i\theta_i + a_i\Theta_{ij}a_j)(\gamma + 2a_k\theta_k + a_k\Theta_{k\ell}a_\ell)\} \\ &= E\{\gamma^2\} + 4E\{\theta_i\gamma\}a_i + 2E\{\Theta_{ij}\gamma\}a_ia_j + 4E\{\theta_i\theta_k\}a_ia_k \\ &\quad + 4E\{\Theta_{ij}\theta_k\}a_ia_ja_k + E\{\Theta_{ij}\Theta_{k\ell}\}a_ia_ja_ka_\ell , \end{aligned}$$

which, by (8-25) and (8-28), becomes

$$\begin{aligned} E\{\mu^2\} &= \alpha^2 + 4\alpha^2\delta_{ik}a_ia_k + \alpha^2\delta_{ik}\delta_{j\ell}a_ia_ja_ka_\ell \\ &= \alpha^2(1 + 4a_ia_i + a_ia_ia_ka_k) \end{aligned}$$

or $$E\{\mu^2\} = \alpha^2\left[1 + a^Ta(4 + a^Ta)\right] .$$

After back-substitution for a we finally get

$$E\{\mu^2\} = \alpha^2\left[1 + C_P^TC^{-1}C^{-1}C_P(4 + C_P^TC^{-1}C^{-1}C_P)\right] . \tag{8-29}$$

The variance α^2 of the deviation of an element of $\tilde{C}$ or $\tilde{C}_P$ from the true values can be estimated as soon as the problem is defined and the approximation model is chosen (see chapters 2-5). For small α^2, C_P and C can be replaced by $\tilde{C}_P$ and $\tilde{C}$, respectively, so that equations (8-26) and (8-29) provide immediately an estimate of the consequences of the covariance function approximation. Although the stochastic model discussed here does not reflect the physical reality, it can be very useful and gives, at least for large scale applications, not such pessimistic error estimates as we obtain with norm estimates. The norm of δC_P and δC is, so to say, replaced by the global value α^2.

9. Approximation of the spatial covariance function by a bicubic spline function

So far all our approximation models discussed here were restricted to the sphere or plane only. Since geodetic measurements are performed at points on and outside the surface of the earth we need a spatial covariance function in order to be able to combine heterogeneous data in a consistent way. The basic properties of the covariance function have already been discussed in chapter 2 where we have pointed out that the homogeneous and isotropic covariance function harmonic outside some internal sphere depends essentially on two variables only, the product $r_P r_Q$ and the spherical distance ψ. This fact will be used here for an approximation of the covariance function by a bicubic spline function.

9.1 The bicubic spline function

A short definition of a cubic spline was given at the beginning of chapter 5. The bicubic spline is just an extension to two dimensions: A function consisting of bicubic polynomials

defined on a regular grid is called a bicubic spline function if it is twice continuously differentiable with respect to each independent variable over the whole domain and if it fulfills certain prescribed boundary conditions on the boundary of the domain. If this spline interpolates data it is called an interpolating bicubic spline. The coefficients of each bicubic polynomial are uniquely determined by continuity conditions up to and including the second order derivative together with known boundary conditions. Within each subrectangular area of the mesh its equation is given by

$$f^{(ij)}(\rho,\psi) = \sum_{k=0}^{3}\sum_{\ell=0}^{3} a_{k\ell}^{(ij)} (\rho-\rho_i)^k (\psi-\psi_j)^\ell \quad . \tag{9-1}$$

The coefficients $\{a_{k\ell}^{(ij)}\}$ are the result of a matrix product (Meissl, 1971, p.36)

$$A^{(ij)} := \{a_{k\ell}^{(ij)}\} = H^T(g)\,\Gamma\,H(h)$$

with

$$H(h) = \begin{bmatrix} 1 & 0 & -3/h^2 & 2/h^3 \\ 0 & 1 & -2/h & 1/h^2 \\ 0 & 0 & 3/h^2 & -2/h^3 \\ 0 & 0 & -1/h & 1/h^2 \end{bmatrix} , \tag{9-2a}$$

h ... grid distance in ρ-direction,
g ... grid distance in ψ-direction,

and

$$F = \begin{bmatrix} f_{ij} & q_{ij} & f_{i,j+1} & q_{i,j+1} \\ p_{ij} & r_{ij} & p_{i,j+1} & r_{i,j+1} \\ f_{i+1,j} & q_{i+1,j} & f_{i+1,j+1} & q_{i+1,j+1} \\ p_{i+1,j} & r_{i+1,j} & p_{i+1,j+1} & r_{i+1,j+1} \end{bmatrix} \tag{9-2b}$$

f_{ij} ... function values at the gridpoint with the indices (ij)
p_{ij} ... first derivative with respect to ρ at (ij)
q_{ij} ... first derivative with respect to ψ at (ij)
r_{ij} ... second derivative with respect to $\rho\psi$ at (ij) .

The first and second derivatives at the mesh points are determined by continuity conditions similar to the one-dimensional case. For a unique representation, however, these values must be known at the boundary points and are, therefore, called boundary values. Usually these boundary values are not known and, consequently, some assumptions concerning their values have to be made. In our case, however, we are in the lucky position that we know all these boundary values exactly (they are just derivatives of the covariance function) and, therefore, the spline representation of the spatial covariance function is unique.

9.2 Approximated covariance expressions

The spatial covariance function of the disturbing potential

$$K(P,Q) = \sum_{n=0}^{\infty} k_n \left(\frac{R_b^2}{r_P r_Q}\right)^{n+1} P_n(\cos\psi)$$

will now be approximated by a bicubic spline function (9-1). As independent variables we will choose

$$\rho := \frac{R_b^2}{r_P r_Q}$$

and $$\psi = \arccos \frac{\underline{r}_P \cdot \underline{r}_Q}{|\underline{r}_P|\,|\underline{r}_Q|}$$

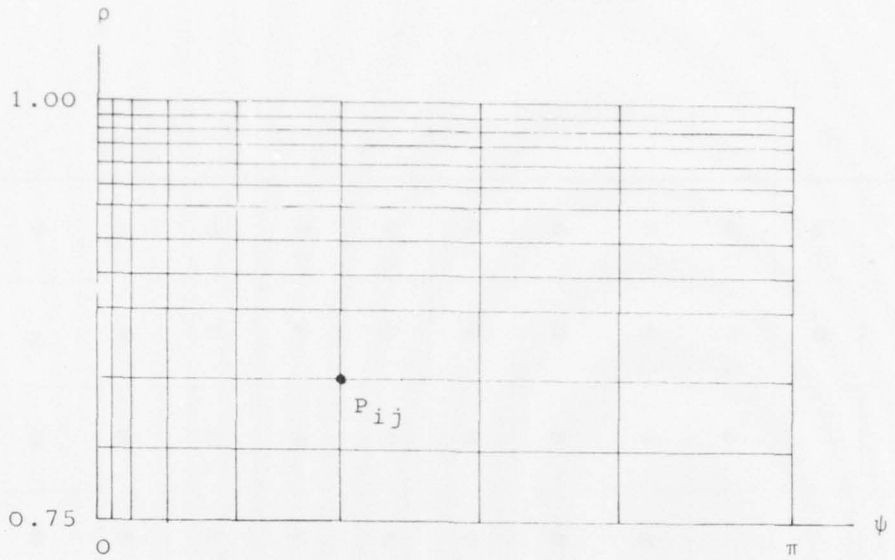


Fig. 9.1 Possible bicubic spline gridding

Since the bicubic spline is twice continuously differentiable with respect to each independent variable ρ and ψ, its third derivative with respect to one variable is a step function. Consequently, at most 3 differentiations with respect to each variable ρ and ψ are admitted. With this in mind we summarize in table 1 all geodetically relevant covariance expressions which can be derived from an approximation of the disturbing potential covariance function by a bicubic spline function. (Recall that a radial derivative corresponds to a differentiation with respect to ρ, a horizontal derivative to a differentiation with respect to ψ.) All covariances denoted by ● can be derived from the spline representation of the disturbing potential covariance function. A bilinear representation, for example, would only permit expressions denoted by ○, a step function representation only $cov(T,T)$. Hatched areas indicate covariances which cannot directly be derived from the basic spline representation. It is, however possible, also to fill up these gaps by

	T	Δg	ξ	η	$\frac{\partial \Delta g}{\partial \theta}$	$\frac{\partial \Delta g}{\partial \lambda}$	$\frac{\partial^2 T}{\partial r \partial \theta}$	$\frac{\partial^2 T}{\partial r \partial \lambda}$	$\frac{\partial^2 T}{\partial \theta^2}$	$\frac{\partial^2 T}{\partial \theta \partial \lambda}$	$\frac{\partial^2 T}{\partial \lambda^2}$	$\frac{\partial^2 T}{\partial r^2}$
T	● ○	● ○	● ○	● ○	● ○	● ○	● ○	● ○	●	●	●	●
Δg		●	● ○	● ○	●	●	●	●	●	●	●	●
ξ			●	●	●	●	●	●	●	●	●	●
η				●	●	●	●	●	●	●	●	●
$\frac{\partial \Delta g}{\partial \theta}$					●	●	●	●	●	●	●	●
$\frac{\partial \Delta g}{\partial \lambda}$						●	●	●	●	●	●	●
$\frac{\partial^2 T}{\partial r \partial \theta}$							●	●	●	●	●	●
$\frac{\partial^2 T}{\partial r \partial \lambda}$								●	●	●	●	●
$\frac{\partial^2 T}{\partial \theta^2}$												●
$\frac{\partial^2 T}{\partial \theta \partial \lambda}$												●
$\frac{\partial^2 T}{\partial \lambda^2}$												●
$\frac{\partial^2 T}{\partial r^2}$												

Table 1 Possible covariance approximations

●... spline

○... linear

choosing additional spline representations which interpolate the radial and horizontal derivative.

Until now we have only stated that the different covariance expressions listed in table 1 can be derived by simple differentiations of the spline; we suppressed the fact that besides these differentiations there are also partial derivatives of ρ with respect to r_P and r_Q and partial derivatives of ψ with respect to θ_P, θ_Q, λ_P and λ_Q involved. In the sequel we will give a setting-up of some frequently used covariances. Details concerning the derivations can be found in (Tscherning, 1976), the relevant relations between different quantities of the gravity field can be found in (Heiskanen and Moritz, 1967).

$$\mathrm{cov}(T_P, T_Q) = f$$

$$\mathrm{cov}(T_P, \Delta g_Q) = f_\rho \frac{\rho}{r_Q} - f\frac{2}{r_Q}$$

$$\mathrm{cov}(T_P, \xi_Q) = \frac{1}{\gamma r_Q} f_\psi \psi_{\theta_Q}$$

$$\mathrm{cov}(T_P, \eta_Q) = -\frac{1}{\gamma r_Q \sin\theta_Q} f_\psi \psi_{\lambda_Q}$$

$$\mathrm{cov}(\Delta g_P, \Delta g_Q) = \frac{1}{r_P r_Q}(f_{\rho\rho}\rho^2 - 3f_\rho\rho + 4f) \qquad (9\text{-}4)$$

$$\mathrm{cov}(\Delta g_P, \xi_Q) = \frac{\psi_{\theta_Q}}{\gamma r_P r_Q}(f_{\psi\rho}\rho - 2f_\psi)$$

$$\mathrm{cov}(\Delta g_P, \eta_Q) = \frac{-\psi_{\lambda_Q}}{\gamma r_P r_Q \sin\theta_Q}(f_{\psi\rho}\rho - 2f_\psi)$$

$$\mathrm{cov}(\xi_P, \xi_Q) = \frac{1}{\gamma^2 r_P r_Q}(f_{\psi\psi}\psi_{\theta_P}\psi_{\theta_Q} + f_\psi \psi_{\theta_P\theta_Q})$$

$$\mathrm{cov}(\xi_P, \eta_Q) = \frac{-1}{\gamma^2 r_P r_Q \sin\theta_Q}(f_{\psi\psi}\psi_{\theta_P}\psi_{\lambda_Q} + f_\psi \psi_{\theta_P\lambda_Q})$$

$$\mathrm{cov}(\eta_P,\eta_Q) = \frac{1}{\gamma^2 r_P r_Q \sin\theta_P \sin\theta_Q}\left(f_{\psi\psi}\psi_{\lambda_P}\psi_{\lambda_Q} + f_{\psi}\psi_{\lambda_P\lambda_Q}\right) .$$

The partial derivatives of the spherical distance with respect to the spherical coordinates of P and Q can easily be derived from

$$\psi_{(\cdot)} = \frac{\partial\psi}{\partial(\cdot)} = -\frac{1}{\sin\psi}\frac{\partial\cos\psi}{\partial(\cdot)}$$

with $\cos\psi = \cos\theta_P\cos\theta_Q + \sin\theta_P\sin\theta_Q\cos(\lambda_Q - \lambda_P)$
All derivatives of this kind can be found in (Tscherning, 1976, pp.18,19). Also very simple is the calculation of partial derivatives of the spline with respect to ρ and ψ. Recalling its defining equation (9- 1), we obtain by primitive differentiation the partial derivatives occuring in the covariance expressions above:

$$f_{\rho}^{(ij)}(\rho,\psi) = \sum_{k=1}^{3} k \sum_{\ell=0}^{3} a_{k\ell}^{(ij)}(\rho-\rho_i)^{k-1}(\psi-\psi_j)^{\ell}$$

$$f_{\rho\rho}^{(ij)}(\rho,\psi) = \sum_{k=2}^{3} k(k-1) \sum_{\ell=0}^{3} a_{k\ell}^{(ij)}(\rho-\rho_i)^{k-2}(\psi-\psi_j)^{\ell}$$

$$f_{\psi}^{(ij)}(\rho,\psi) = \sum_{k=0}^{3}\sum_{\ell=1}^{3} \ell a_{k\ell}^{(ij)}(\rho-\rho_i)^{k}(\psi-\psi_j)^{\ell-1} \qquad (9\text{-}5)$$

$$f_{\psi\psi}^{(ij)}(\rho,\psi) = \sum_{k=0}^{3}\sum_{\ell=2}^{3} \ell(\ell-1) a_{k\ell}^{(ij)}(\rho-\rho_i)^{k}(\psi-\psi_j)^{\ell-2}$$

$$f_{\psi\rho}^{(ij)}(\rho,\psi) = \sum_{k=1}^{3} k \sum_{\ell=1}^{3} \ell a_{k\ell}^{(ij)}(\rho-\rho_i)^{k-1}(\psi-\psi_j)^{\ell-1}$$

In this way all derivatives up to $f_{\rho\rho\rho\psi\psi\psi} = 36a_{33}$ can be calculated

9.2 How to get covariances

At the very beginning of this report we have mentioned that the covariance function approximations can be useful for large scale applications of collocation. Small scale problems are of no concern; it would be a waste of time to perform the intermediate step of the approximation. Also we want to avoid the impression that this report is an attack on the excellent and extremely useful subroutine COVAX written by C.C. Tscherning. This sophisticated subroutine, however, should only be used as a mean for obtaining an accurate network of absolute covariances. Inside the network we can perform, so to say, relative calculations called interpolations based on a much simpler function compared to the covariance function. (No geodesist will use a first order theodolite for purposes of property dividing!) In this sense covariance function approximations have to be understood.

We will now write down in a compact form all steps necessary for A) setting up the spline representation and B) calculating different kinds of covariances.

A) 1. Select an appropriate grid spacing in ρ- and ψ-direction (g and h) according to some kind of approximation error estimates.

2. Define the lower and upper bounds in ρ and ψ you are interested in and generate a rectangular mesh ($i = 1,\ldots, I$; $j = 1,\ldots, J$).

3. Calculate the covariances of the disturbing potential at all mesh points using the subroutine COVAX.

4. Calculate the boundary values necessary for the spline representation:

$$\left.\frac{\partial K}{\partial \psi}\right|_{\substack{\rho=\rho_i \\ \psi=0}} = q_{i1} = 0$$ because of the symmetry of K with respect to $\psi = 0$.

$$\left.\frac{\partial K}{\partial \psi}\right|_{\substack{\rho=\rho_i \\ \psi=\psi_J}} = q_{iJ} \neq 0 \quad \text{in general}$$

$$\left.\frac{\partial K}{\partial \rho}\right|_{\substack{\rho=\rho_1 \\ \psi=\psi_j}} = p_{1j} \neq 0 \quad \text{in general}$$

$$\left.\frac{\partial K}{\partial \rho}\right|_{\substack{\rho=\rho_I \\ \psi=\psi_j}} = p_{Ij} \neq 0 \quad \text{in general}$$

$$\left.\frac{\partial^2 K}{\partial \rho\, \partial \psi}\right|_{\substack{\rho=\rho_1 \\ \psi=O^1}} = r_{11} = 0$$

$$\left.\frac{\partial^2 K}{\partial \rho\, \partial \psi}\right|_{\substack{\rho=\rho_I \\ \psi=O^I}} = r_{I1} = 0$$

$$\left.\frac{\partial^2 K}{\partial \rho\, \partial \psi}\right|_{\substack{\rho=\rho_1 \\ \psi=\psi_J}} = r_{1J} \neq 0 \quad \text{in general}$$

$$\left.\frac{\partial^2 K}{\partial \rho\, \partial \psi}\right|_{\substack{\rho=\rho_I \\ \psi=\psi_J}} = r_{IJ} \neq 0 \quad \text{in general.}$$

5. With the covariances at the mesh points and the boundary values defined above calculate by some spline algorithm all other first and second mixed derivatives; calculate the spline coefficients $\{a_{k\ell}^{(ij)}\}$ for all subrectangles. Store these coefficients on a permanent file.

From now on the calculation of covariances is extremely simple:

B) 1. Calculate ρ and ψ and find the corresponding subrectangle (ij) such that $\rho_i \leq \rho < \rho_{i+1}$, $\psi_j \leq \psi < \psi_{j+1}$.
2. Define the kind of covariances you are interested in and calculate it according to formulas of the kind of (9-4) and (9-5).

What are the advantages of the spline representation compared to the bilinear or step function representation?

- The spline is a very accurate interpolating element; therefore, the number of grid points can be kept small and we save a lot of mass-storage.
- The spline representation admits a great number of different kinds of covariances derived from the basic spline; again we save mass-storage.

However, there are also some disadvantages:

- The calculation of the spline coefficients is more laborious than for the other elements.
- The calculation of a covariance needs more calculations and therefore more CPU-time. These calculations, however, consist of very simple operations only so that the routine is still extremely fast compared to exact covariance calculations.

The following table 2 gives a comparison of CPU-time needed for the calculation of some kinds of covariances using the subroutine COVAX on the one hand and the spline representation on the other hand

	T	Δg	ξ	η	
T	24.0	25.3	38.7	38.6	COVAX
	(0.4)	(0.9)	(0.6)	(0.6)	Spline
Δg		26.3	39.3	39.9	
		(1.1)	(0.5)	(0.5)	
ξ			54.3	53.3	
			(0.6)	(1.0)	
η				53.6	
				(1.0)	

Table 2: CPU-time used for 1000 calculations of covariances using COVAX and a spline representation; computer: UNIVAC 494.

10. Conclusions

Recently, many new sources of geodetic data are becoming available, all of them being related somehow to the gravity field of the earth. The least-squares collocation method is the mathematical tool for handling all these heterogeneous data in a consistent way. The more sophisticated the kind of measurement is, the more expensive is its processing. A typical example is the use of satellite-to-satellite tracking data where the covariances are obtained by multiple integration procedures with, again, covariances as input. Since the exact calculation of covariances is fairly time consuming and a large number of tracking data is to be expected, such an application for geodetic purposes would be extremely expensive. This fact made the question arise whether it is possible to avoid frequent calls of the highly sophisticated subroutine COVAX by using other simple functions approximating the covariance function.

The basic principle underlying these investigations is well known and frequently applied in many fields; the network principle: generate a net of fixed points (here grid points) and perform very accurate measurements at these points (here, calculate exact covariances); these fixed points serve as a basis for small scale measurements which can be performed using simpler apparatus (here, more or less interpolation of covariances by means of finite elements). This report was primarily devoted to the study of interpolation errors, perturbation of spectra and to the consequences of the approximation for the predicted signal and its mean square error. Because of its smoothness and its most favourable approximation properties the spline function representation of the covariance function presents itself as a very useful tool for this kind of application.

Acknowledgement

The author wishes to thank Prof.Dr. H. Moritz for many valuable discussions. The preparation of computer programs concerning these investigations has been partly sponsored by the "Österreichischer Fond zur Förderung der wissenschaftlichen Forschung". Computer time has been made available by the "Rechenzentrum Graz".

References

Ahlberg, J.H., E.N. Nilson and J.L. Walsh (1967): The Theory of Splines and Their Applications. Academic Press, London-New York.

Davis, P.J. (1975): Interpolation and Approximation. Dover Publications, Inc., New York.

Friedrichs, K.O. (1973): Spectral Theory of Operators in Hilbert Space. Springer-Verlag, New York.

Gradshteyn, I.S. and I.W. Ryzhik (1965): Table of Integrals, Series, and Products. Academic Press, New York.

Heiskanen, W.A. and H. Moritz (1967): Physical Geodesy. W.H. Freeman & Company, San Francisco.

Hobson, E.W. (1965): The Theory of Spherical and Ellipsoidal Harmonics. Chelsea Publishing Company, New York.

Kryński, J. (1978): Possibilities of low-low satellite-to-satellite tracking for local geoid improvement. Mitteilungen der geodtäischen Institute der TU Graz, Folge 31.

Lebedev, N.N. (1965): Special Functions and Their Applications. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Meissl, P. (1971): Preparations for the numerical evaluation of second order Molodensky-type formulas. Report No. 163, Department of Geodetic Science, The Ohio State University, Columbus.

Moritz, H. (1970): Least-squares estimation in physical geodesy. Report No. 130, Department of Geodetic Science, The Ohio State University, Columbus.

Moritz, H. (1976): Covariance functions in least-squares collocation. Report No. 240, Department of Geodetic Science, The Ohio State University, Columbus.

Papoulis, A. (1968): Systems and Transforms with Applications in Optics. McGraw-Hill, New York.

Rapp, R.H. and C.C. Tscherning (1974): Closed covariance expressions for gravity anomalies, geoid undulations and deflections of the vertical implied by anomaly degree variance models. Report No. 208, Department of Geodetic Science, The Ohio State University, Columbus.

Schwarz, K.P. (1976): Geodetic accuracies obtainable from measurements of first and second order gravitational gradients. Report No. 242, Department of Geodetic Science, The Ohio State University, Columbus.

Strang, G. and G.J. Fix (1973): An Analysis of the Finite Element Method. Prentice-Hall, Inc., Englewood cliffs, New Jersey.

Sünkel, H. (1977): Die Darstellung geodätischer Integralformeln durch bikubische Spline-Funktionen. Mitteilungen der geodätischen Institute der TU Graz, Folge 28.

Tscherning, C.C. (1976): Covariance expressions for second and lower order derivatives of the anomalous potential. Report No. 225, Department of Geodetic Science, The Ohio State University, Columbus.

Wilkinson, J.H. and C. Reinsch (1971): Linear Algebra. Springer-Verlag, Berlin.